Brigham Young University
BYU ScholarsArchive

All Theses and Dissertations

2011-08-03

The Characterization of Avian Polyomavirus, Satellite Tobacco Mosaic Virus, and Bacteriophage CW02 by Means of Cryogenic Electron Microscopy

Peter S. Shen
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

 Part of the Biochemistry Commons, and the Chemistry Commons

BYU ScholarsArchive Citation

Shen, Peter S., "The Characterization of Avian Polyomavirus, Satellite Tobacco Mosaic Virus, and Bacteriophage CW02 by Means of Cryogenic Electron Microscopy" (2011). *All Theses and Dissertations*. 3069.
https://scholarsarchive.byu.edu/etd/3069

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

The Characterization of Avian Polyomavirus, Satellite Tobacco Mosaic Virus, and Bacteriophage CW02 by Means of Cryogenic Electron Microscopy

Peter S. Shen

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

David M. Belnap, chair
Bradford K. Berges
Gregory F. Burton
Allen R. Buskirk
Keith A. Crandall

Department of Chemistry and Biochemistry

Brigham Young University

December 2011

Copyright © 2011 Peter S. Shen

All Rights Reserved

ABSTRACT

The Characterization of Avian Polyomavirus, Satellite Tobacco Mosaic Virus, and Bacteriophage CW02 by Means of Cryogenic Electron Microscopy

Peter S. Shen
Department of Chemistry and Biochemistry, BYU
Doctor of Philosophy

Viruses are the most abundant biological entity in the biosphere and are known to infect hosts from all domains of life. The aim of my work is to identify conserved and non-conserved features among the capsid structures of related and divergent icosahedral viruses via cryogenic electron microscopy, sequence analysis, molecular modeling, and other techniques.

Bird polyomaviruses often cause severe disease in their hosts whereas mammalian polyomaviruses generally do not. Avian polyomavirus is a type of bird polyomavirus with an unusually broad host range compared to the restricted tropism of other polyomaviruses. Although most polyomaviruses have a conserved, rigid capsid protein structure, avian polyomavirus has a flexible capsid shell and a non-conserved C-terminus in its major capsid protein. A β-hairpin motif appears to stabilize other polyomaviruses but is missing in avian polyomavirus. The lack of this structure in avian polyomavirus may account for its capsid flexibility and broad host range. A minor capsid protein unique to bird polyomaviruses may be located on the inner capsid surface. This protein may have a role in the acute disease caused by bird polyomaviruses.

The solution-state capsid structure of satellite tobacco mosaic virus was unexpectedly different than the previously solved crystalline structure. The conformational differences were accounted for by a shift of the capsid protein about the icosahedral fivefold axis. Conversely, the RNA core was consistent between solution and crystalline structures. The stable RNA core supports previous observations that the viral genome stabilizes the flexible capsid.

Halophage CW02 infects *Salinivibrio* bacteria in the Great Salt Lake. The three-dimensional structure of CW02 revealed a conserved HK97-like fold that is found in all tailed, double-stranded DNA viruses. The capsid sequence of CW02 shares less than 20% identity with HK97-like viruses, demonstrating that structure is more conserved than sequence. A conserved module of genes places CW02 in the viral T7 supergroup, members of which are found in diverse aquatic environments. No tail structure was observed in reconstructions of CW02, but turret-like densities were found on each icosahedral vertex, which may represent unique adaptations similar to those seen in other extremophilic viruses.

Keywords: cryogenic electron microscopy, halophage, icosahedral virus, polyomavirus, satellite tobacco mosaic virus, single-particle reconstruction, structural evolution

Acknowledgments

The challenges I faced throughout my studies were met with immeasurable support and guidance. I am privileged to have David Belnap as my adviser and mentor, who has been the sounding board in each of my projects. I thank my committee members for their constructive criticisms and their unique, broad perspectives. I am grateful to members of the Belnap lab for their diligence and camaraderie. In particular, I would like to recognize Eduardo Sanz-García for his technical support within the lab. Contributors to my projects are acknowledged at the end of their respective chapters.

Funding for my projects was provided by BYU institutional funds; by fellowships from the BYU Department of Chemistry and Biochemistry, the BYU Cancer Research Center, and BYU Graduate Studies; and by a grant from the National Institute of Allergy and Infectious Disease.

Finally, I thank my family for their continual, unconditional support; my wife, Jenny, for her tolerating my long hours away from home, ensuring my constant nourishment, and encouraging me when I despair; my sons, Elijah and Noah, for keeping me grounded in appreciating the simple things in life; I also thank my parents for their emotional support and for their examples in living life to the fullest each day.

"...by the help of Microscopes, there is nothing so small, as to escape our inquiry; hence there is a new visible World discovered to the understanding."

—Robert Hooke, *Micrographia* (1665) (sic)

Table of Contents

List of Figures

Chapter 1

Introduction to virus structure

1.1 Overview

Viruses are the most abundant biological entity on earth. Since the initial characterization of tobacco mosaic virus by Martinus Beijerinck in 1898, viruses have been found in nearly all ecosystems and infect organisms from all domains of life. Recent estimations place the number of virus particles in the biosphere on the order of 10^{31}-10^{32}, outnumbering host cells by at least 10 to 1 (*1*). The prevalence of virus activity also makes viruses a dominant contributor to selective pressure on host organisms (*2*). As host organisms adapt to the constant threat of virus infection, viruses engage in metaphorical arms races by continually developing countermeasures to overcome host defenses (*3*).

Human immune systems have evolved to thwart infection, and only a minute fraction of viruses significantly impact human health. Yet, viruses account for many lethal pandemics throughout history and are a persistent threat to human populations due to constant evolution. For instance, the emergence of HIV and SARS are recent pandemics exacerbated by virus evolution while other deadly viruses such as Ebola and Marburg viruses have pandemic potential. The role of viruses in human evolution is demonstrated by the presence of endogenous retroviral elements comprising up to 8% of human genomes (*4*). Today, viruses are estimated to cause between 10-15% of human cancers (*5*). Virus outbreaks in wildlife and agricultural settings, such as those caused by avian influenza and foot-and-mouth disease virus, also affect human well-being.

1.2 Virus composition

Viruses are small particles, typically ranging from tens of nanometers to several hundred in diameter. All viruses consist of a protein coat that surrounds the viral genome. In 1957, Watson and Crick pointed out that viruses carrying "limited amount of [nucleic acid] can only carry a limited amount of information. Thus the protein molecules of the virus can only be of limited size" (*6*). They postulated that virus coats (capsids) are constructed from identical subunits. High-resolution structures of virus capsids and genomic analysis has since confirmed that many viral genomes encode at least one major capsid protein, which serves as a repetitive building block that is combined into an enclosed, symmetrical structure. Various modes of assembly of the major capsid proteins give rise to diverse virus shapes (morphologies), most notably icosahedral and helical structures.

Some virus genomes encode minor capsid proteins that accompany the major capsid proteins. As with major capsid proteins, there are several copies of minor capsid proteins per virus particle. The functions of minor capsid proteins are diverse—some aid in the cell infection process, some aid in stabilizing the virus structure, and so forth.

All viruses are built so that their encapsulated genomes can be transmitted into host cells. The foremost function of virus capsids is to protect the viral genome. Capsids must withstand potential chemical and physical obstacles in their surrounding environments, including proteolytic or nucleolytic enzymes and pH changes. When a virus particle encounters a host cell, the infection mechanism is initiated by recognition and interaction with a host molecule. Viruses must also have a means of delivering the encased genome upon reaching a host cell, which generally requires a series of conformational changes in the virus capsid. Thus, virus capsids

must be sufficiently robust to withstand extracellular environments, but also sufficiently pliable to allow for the timely release of their genomes.

The composition of virus genomes is another diverse feature among viruses. Most virus genomes are either composed of DNA or RNA and are either single-stranded (ss) or double-stranded (ds). The polarity, or sense, of ssRNA genomes adds an additional layer of diversity among viruses. Some viruses (e.g., influenza) carry an antisense (negative-sense) ssRNA genome and require a virus-encoded RNA-dependent RNA polymerase for progeny production. Other viruses (e.g., poliovirus) carry positive-sense ssRNA genes that can be directly accessed by host ribosomes to produce proteins. Further, viral genomes can be either circular or linear. Genome sizes also vary greatly, ranging anywhere between 1 to 1,200 kilobases. High-resolution structures of viruses also show that some viral genomes play a role in maintaining capsid integrity (*7*).

Many viruses also possess a lipid envelope, derived from the host membrane. The envelope adds an additional shield to protect the viral genome, while also embedding a layer of viral glycoproteins that function in recognizing host receptor sites. Upon attachment to the receptor protein, the viral envelope fuses with the host membrane and allows viral entry into the cell.

1.3 Icosahedral symmetry in viruses

Many capsids assemble into an icosahedral lattice, which is defined by 12 vertices with fivefold rotational symmetry, 20 equilateral triangular faces, and 30 edges. The icosahedral geometry is arranged so that each vertex interacts with five of the triangles. Thus, the simplest icosahedral viruses have at least 60 identical proteins that form 12 pentamers (one pentamer per vertex).

Because each protein occupies the same structural environment, the chemical interactions among all capsid proteins are equivalent. This manner of assembly also limits the size of the virus particle, which is dictated by the size of the capsid protein.

In order to accommodate larger genomes, many icosahedral viruses adopt a more complex structure and consist of more than 60 copies of the capsid protein (Fig. 1.1). In such viruses, the icosahedral shell is expanded by incorporating hexamers throughout the icosahedral shell. In 1962, Caspar and Klug (*8*) applied the icosahedral triangulation numbering system to explain how larger icosahedral structures can be formed (Fig. 1.1). The triangulation (*T*) number of any icosahedron is the sum of $h^2 + hk + k^2$, where the integers *h* and *k* describe the spacing between pentamers. *T* numbers also correspond to the number of smaller equilateral triangles formed by the subdivision of one icosahedral face (Fig. 1.1). In general, the total number of subunits is 60*T*.

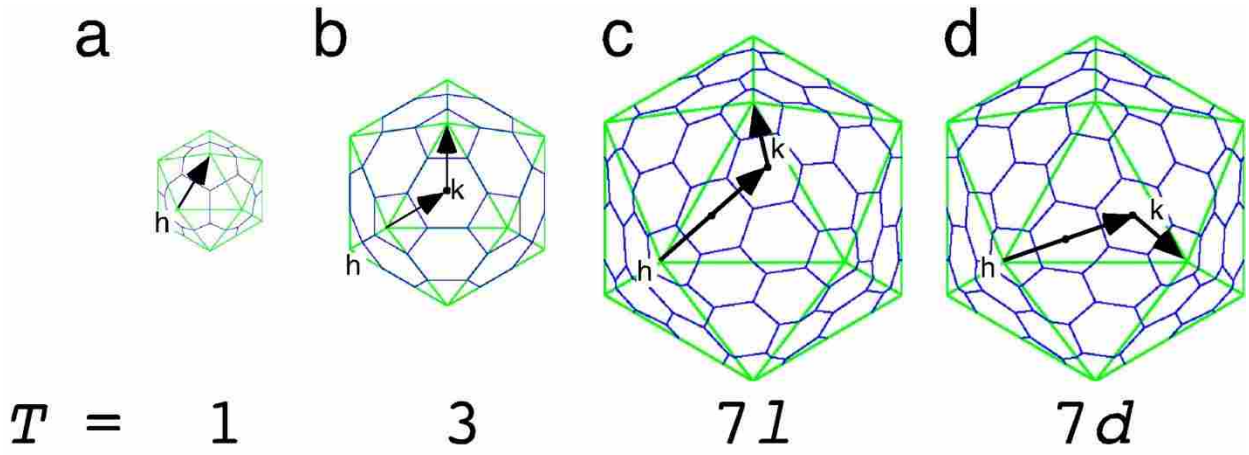


Figure 1.1. Icosahedral lattices and T numbers. Viruses may adopt icosahedral structures of varying complexities (blue) but maintain the same overall shell (green). Vertices of the subdivided triangles are the centers of hexagons (dots). Structures are shown to scale. (a) The simplest icosahedron is comprised of 12 pentamers, where h = 1 and k = 0. (b) The addition of a hexamer at each face results in greater spacing between pentamers, increasing the value of k from 0 to 1. (c and d) The greater the number of hexamers, the greater the spacing between pentamers, thus increasing the values of h and k. In this case, an additional hexamer is added in the h direction, resulting in the values h = 2 and k = 1. If h and k are both nonzero and $h \neq k$, then the resulting icosahedron will have a non-equivalent enantiomer, designated as either laevo (l, left-handed) or dextro (d, right-handed). (Icosahedral projection files obtained from VIPERdb [http://viperdb.scripps.edu])

The triangulation numbering system applied by Caspar and Klug also introduced the principle that virus particles with more than 60 copies of the capsid protein cannot occupy structurally identical positions. In other words, because some proteins must occupy the fivefold vertices and others must occupy the interwoven hexagons, the proteins must therefore occupy "quasiequivalent"—similar, but non-identical—environments. For example, 60 of the 420 subunits of a typical $T = 7$ virus are retained at the 12 fivefold vertices, but the remaining 360 subunits form hexamers and occupy similar, but quasiequivalent, positions (Fig. 1.1).

Structures of some viruses showed radical departures from the constraints imposed by quasiequivalence. For instance, polyomaviruses (described in Chapter 3) consist of 360 subunits, corresponding to a $T = 6$ triangulation number if one uses the formula $T = (\text{subunits}/60)$. Yet, a triangulation number of 6 is impossible by the $h^2 + hk + k^2$ formula. Structural studies of polyomavirus capsids revealed that it consists entirely of pentamers to form the $T = 7d$ lattice (*9*). Also underscoring the diversity of icosahedral virus assembly was the discovery that, although it contains pentamers and hexamers, the capsid of bacteriophage HK97 uses covalent cross-links to form a protein catenane that stabilizes the capsid (see Chapter 5) (*10*).

1.4 Structural evolution of viruses

The evolutionary origins of viruses remain a highly debated subject, complicated by the fact that fossils of viruses have not been observed. Thus, inferences of virus evolution are limited to the study of viruses isolated in the present. Although high evolutionary rates among virus genomes can obscure traces of virus lineage, the structural characterization of some viruses reveals conserved features despite little to no similarity on the sequence level.

1.4.1 The conserved double jellyroll fold

3D structural evolution came to light when the first crystal structures of animal viruses, the human rhinovirus (*11*) and human poliovirus (*12*), showed similar structures with other previously solved icosahedral plant viruses. Thus, the first indication of virus structure spanning across kingdoms of life was established. More recent work has now shown virus structural conservation across domains of life. The protein capsid folds of bacteriophage PRD1 (*13*), human adenovirus (*14*), and archaeal *Sulfolobus* turreted icosahedral virus (STIV) (*15*) resemble one another (Fig. 1.2). The fold of these three viruses, described as a double-jellyroll fold, consists of two separate β-barrel topologies, each composed of eight antiparallel β-strands. The capsid proteins of PRD1, adenovirus, and STIV also have conserved connections between subunits. Similarities among these viruses are not merely limited to the capsid protein fold—they have similar characteristics in their genome organization and mechanism of viral DNA synthesis. Further, these viruses tend to possess distinct proteinaceous spikes on every fivefold vertex. Thus, viruses with common structural elements can be used as a tool for classifying viral lineages.

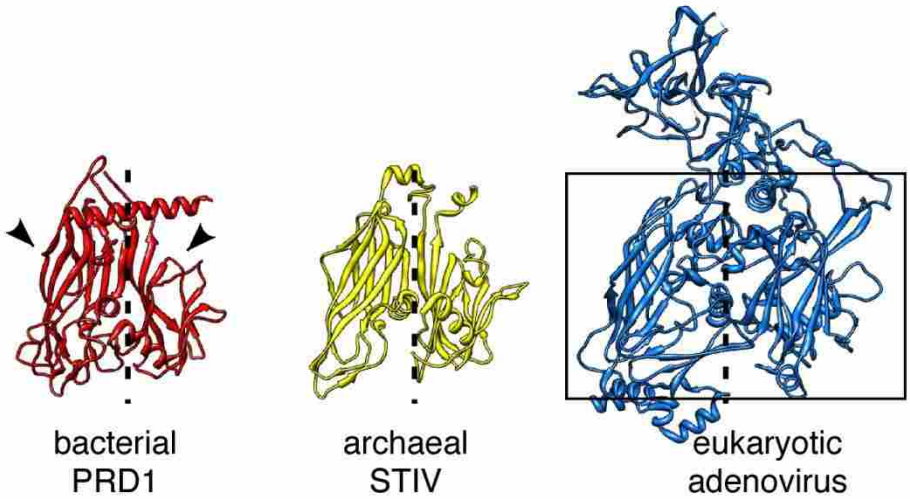


Figure 1.2. The conserved double-jellyroll fold among divergent icosahedral viruses. The views shown here show the division of the two β-barrels down the center of each structure (dashed lines indicate division of , arrowheads point to the two β-barrels of PRD1. Despite the larger structure of adenovirus, the core fold is conserved (boxed). PDB accession IDs: PRD1, 1W8X; STIV, 2BBD; adenovirus, 2OBE.

1.4.2 The HK97-like fold

The observation that divergent viruses share conserved structures led to the idea that viruses may have a more limited number of origins than previously supposed (*16*). This idea is reinforced by the observation that all currently known structures of tailed, dsDNA bacteriophages (order *Caudovirales*) share a conserved fold. The first crystal structure of a tailed, dsDNA phage solved was that of phage HK97 (*17*). High-resolution capsid structures of other tailed, dsDNA phages have since been solved and shown to bear resemblance to that of the HK97 capsid (Fig. 1.3).

The HK97-fold consists of a compact, axial (A) domain and an elongated peripheral (P) domain, which together form a continuous hydrophobic core (*10*). Not only does the fold seem to be conserved in all tailed, dsDNA phages, but the major capsid protein of human herpes simplex virus 1 also adopts the HK97-like fold (*18*). This demonstration provided yet another example that bacteriophages may share a common ancestry with viruses of eukaryotic hosts.

The discovery of viruses with conserved capsid structures spanning all domains of life has supported the hypothesis that viruses previously considered unrelated can be grouped into a reasonably small number of lineages (*14*, *16*). The similarities seen among diverse viruses with the double jellyroll fold and HK97-like fold support this hypothesis. Conversely, the conserved structures could be a product of convergent evolution. Yet, evidence speaks against the latter because viruses with conserved capsid structures also tend to have other similar characteristics such as genomic organization and so forth. Nevertheless, the current opinion in structural virology is that there are only a limited number of protein folds viable for capsid formation. By extension, there are probably only a limited number of assembly mechanisms that give rise to the majority of virus structures.

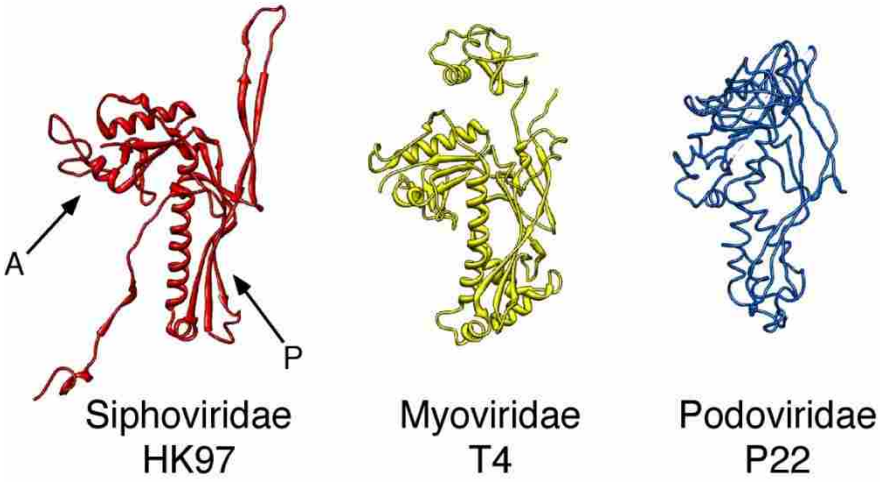


Figure 1.3. The conserved HK97-like fold among the three families of the Caudovirales order (tailed, dsDNA phages). Members of the Siphoviridae family have long, non-contractile tails; myoviruses have long, contractile tails; podoviruses have short, non-contractile tails. Proteins are shown in a similar orientation with A and P domains labeled in HK97. Note that the helical and sheet information is not shown in P22 because it is a pseudo-atomic structure, though the general structure is conserved. PDB accession IDs: HK97, 1OHG; T4, 1YUE; P22, 3IYI.

1.5 Subjects of study

1.5.1 Polyomavirus

Polyomaviruses (PyVs) were first characterized by their ability to induce many (poly-) tumors (-oma) in laboratory mice. Since the discovery of mouse polyomavirus (MPyV) in 1953, at least 23 different polyomaviruses have been found to infect mammals and birds, including five that infect humans (*19*). Despite being genetically similar, mammalian and bird PyVs have different biological characteristics. Mammalian PyVs are generally innocuous towards their hosts, but infections by bird polyomaviruses often cause fatal disease. Mammalian PyVs usually exhibit host- and organ-specific tropism. Conversely, avian polyomavirus (APV) infects many bird species and has been detected in nearly all organs of the infected hosts (*20*). The causes of pathological disparity between mammalian and bird PyVs are unknown.

PyVs are non-enveloped, icosahedral $T = 7d$ viruses and are approximately 50 nm in diameter (Fig. 1.4). Particles contain a circular dsDNA genome of roughly 5,000 base pairs. The capsid is assembled with 72 pentamers, or capsomeres, of the major capsid protein, VP1. Each VP1 subunit within the capsomere contains a core β-barrel jellyroll domain with long C-terminal extensions, or arms (Fig. 1.5A). Twelve of the 72 pentamers occupy the 12 positions of fivefold rotational symmetry, each of which is surrounded by five pentamers (Fig. 1.5B). The remaining 60 pentamers each are surrounded by six other pentamers. The arms span through neighboring pentamers in different manners, which forms the $T = 7$ structure of the capsid. We investigated structural differences between bird and mammalian PyVs and showed that previously unreported inter-capsomere interactions stabilize the capsid (*21*). We also describe non-conserved regions of the C-terminal arm in APV that contribute to capsid flexibility not seen in mammalian PyVs (Fig. 1.4, 1.5C).

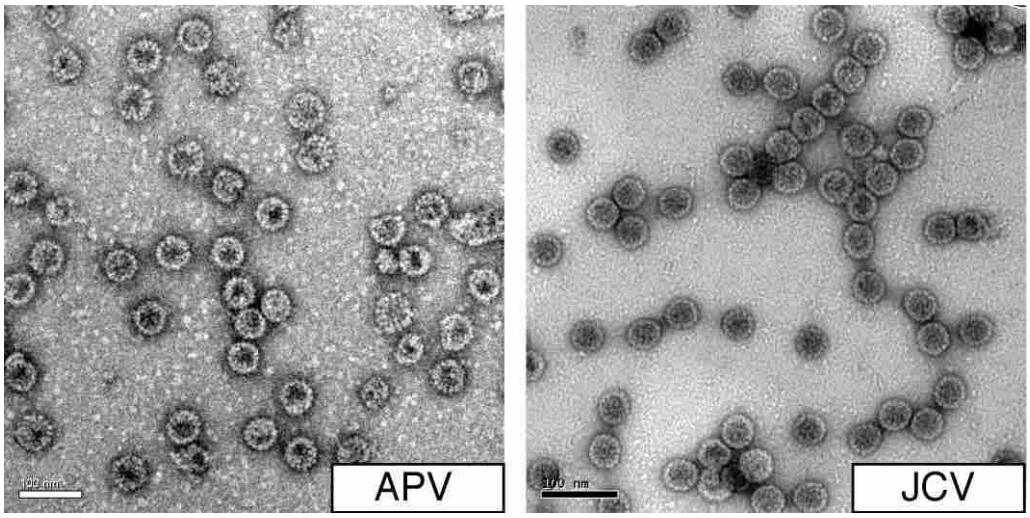


Figure 1.4. Negative stain images of APV and human JC polyomavirus (JCV). Particle sizes of APV are more variable than JCV and other mammalian PyVs.

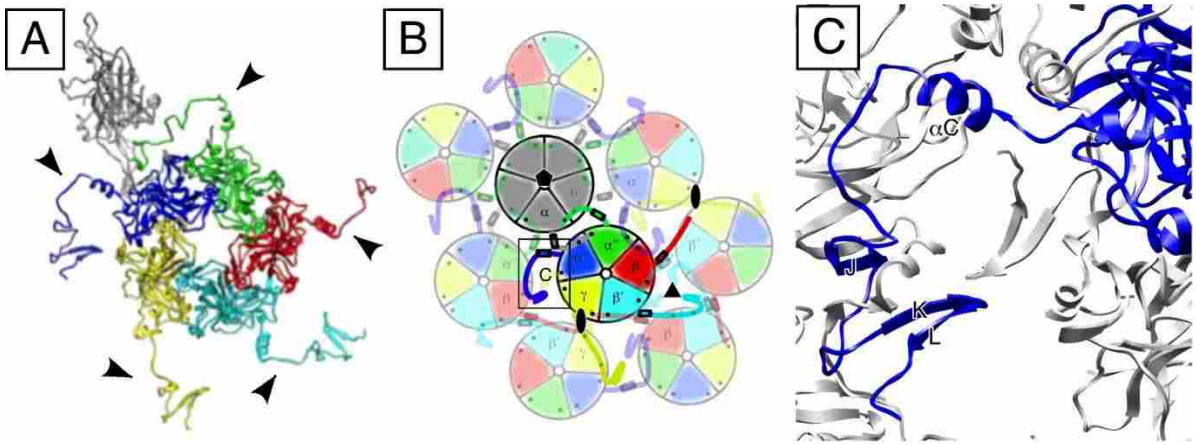


Figure 1.5. The invading 'arms' of polyomaviruses. (A) The six VP1 subunits that comprise the asymmetric unit of PyVs. Emanating C-terminal arms marked by arrowheads. (B) Schematic showing pentavalent (gray) and hexavalent (multi-colored) capsomeres and the manner of 'arm' extension from each VP1. Symmetry axes of the icosahedral capsid shown as pentagon (fivefold), triangle (threefold), and ovals (twofold). Helices shown as cylinders. Calcium binding sites shown as solid dots. Redundant elements faded. (C) Zoomed-in portion of (B) showing the structural components of the MPyV C-terminus (blue).

1.5.2 Satellite tobacco mosaic virus

Satellite tobacco mosaic virus (STMV) is a small, $T = 1$ icosahedral particle approximately 17 nm in diameter (Fig. 1.6A). The capsid is composed of 60 identical copies of the capsid protein and encases a 1,059 nucleotide, positive-sense ssRNA genome (Fig. 1.6B, C). Infection by STMV is only possible through co-infection with a tobamovirus (e.g., TMV) and may heighten symptoms compared to infection by tobamovirus alone. The crystal structure of STMV revealed a regular arrangement of double-helical RNA at every dyad axis (Fig. 1.6C). These 30 helices account for about 50% of the whole genome. The viral genome likely plays an important role in capsid assembly and capsid stability. Extracted RNA cores seem to adopt a globular shape approximately 10 nm in diameter and are able to unfold and fold reversibly (*22*). The goal of our research is to solve the complete structure of the RNA core, but we found previously unreported conformational variation of the capsid, which strengthens the current hypothesis that the RNA genome is important in maintaining capsid structure.

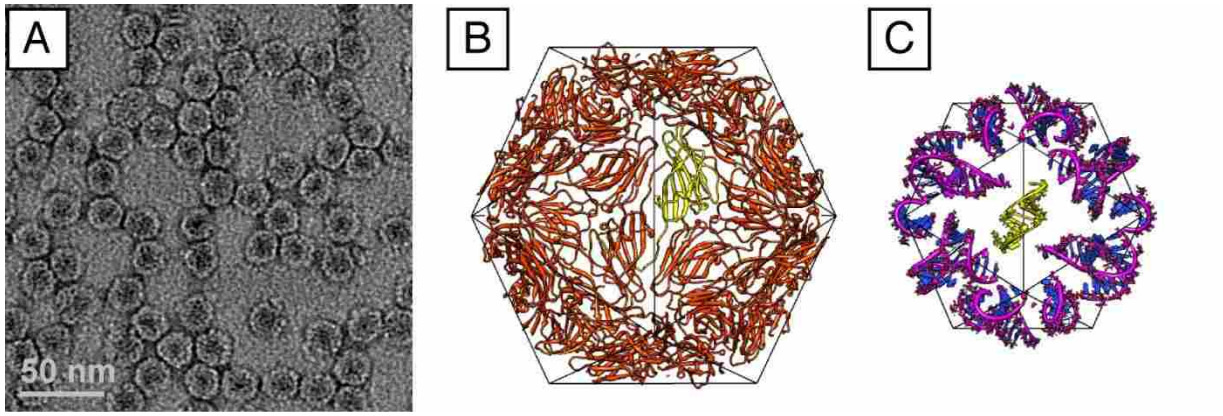


Figure 1.6. The structure of STMV. (A) Electron micrograph of negatively stained electron STMV particles. (B) Atomic-resolution structure of the STMV capsid, revealing the β-jellyroll fold of each of the 60 subunits. (C) Crystallography of STMV particles yielded high-resolution data for approximately 50% of the RNA genome, indicating its rigidity and stability within the particle. Relationship between (B) and (C) shown to scale. A single capsid subunit (B) and RNA duplex (C) is colored in yellow.

1.5.3 Halophage CW02

We solved the cryo-EM structure of bacteriophage CW02, a recently isolated virus from the Great Salt Lake. CW02 (Fig. 1.7), which infects a strain of *Salinivibrio* bacteria, is morphologically similar to other members of the virus family *Podoviridae*, bears a $T = 7l$ capsid, and encapsulates a dsDNA genome. The capsid structure of CW02 is similar to that of HK97-like phages (Fig. 1.2), thus supporting the notion that some viruses evolved from a common viral ancestor. The genomic sequences of HK97-like phages are divergent, indicating that structure is more conserved than sequence.

We used genomic analysis and mass spectrometry to identify the protein components of mature CW02 particles and identified a conserved gene module that encodes the structural proteins of the phage head. The module is conserved with members of the T7 supergroup, which

is divided into two subgroups of phages with (e.g., phage T7) and without (e.g., phage VpV262) an RNA polymerase gene. The CW02 genome does not encode an RNA polymerase. Because CW02 and T7 share similar capsid structures, we conclude that the HK97-like fold predates the divergence of the T7 supergroup into the subgroups with and without an RNA polymerase gene.

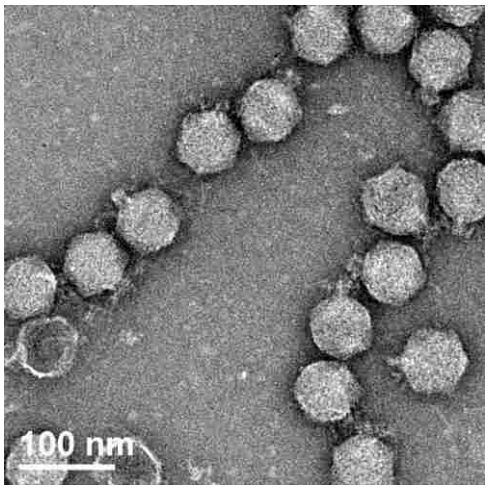


Figure 1.7. Micrograph of negatively stained CW02 particles.

1.6 Biological significance of understanding virus structure

In design, the "form-follows-function" principle states that the shape of an object should be based upon its intended function or purpose. Likewise, this principle holds true on the molecular level: the form or structure of a molecular complex, including viruses, is based upon its function(s). Thus, the study of virus structure is indispensible to understanding their mechanisms of formation (assembly), infection, evolution, and so forth. In fact, the systems of some viruses are so well understood that their activities can be manipulated to achieve other beneficial purposes, such as gene therapy.

Viruses must meet several criteria for successful infection, beginning with recognition and attachment to the host cell. Other components of viral infection include entering or penetrating the host cell, delivering and replicating the viral genome, and assembling and releasing new virus particles. Understanding viral mechanisms can help researchers hone in on specific targets to attenuate their activity. For example, the discovery that papillomaviruses, which are structurally similar to polyomaviruses, could assemble into non-infectious capsids *in vitro* provided the basis for human papillomavirus vaccines. When administered, the vaccine elicits virus-neutralizing antibodies that prevent infection by actual papillomaviruses. A similar strategy has been proposed for APV (*23*).

Virus structures can also shed light on their evolutionary trajectories. Given their short generation times, copious yields of progeny, and high rates of mutation, gene transfer, recombination, and reassortment, the observation that highly divergent viruses adopt conserved protein folds is remarkable. The conserved structures may indicate vulnerable regions that could be targeted by antiviral agents because viruses would presumably have limited adaptability in these regions. For instance, the evolutionary potential of HIV complicates efforts by human immune systems to combat infection. Yet, structures of the HIV envelope glycoproteins gp120 and gp41 revealed regions of conserved, susceptible determinants that could be targeted by vaccine candidates (*24, 25*).

Chapter 2

Overview of methods

2.1 Introduction to cryogenic electron microscopy

The advent of electron microscopy in the mid-1900s marked a turning point in the field of virology. For the first time, viruses could be directly visualized, revealing gross features of their relatively simple structures. Initial methods employed metal-staining techniques that produced images with high contrast and rendered the sample resistant to the high radiation and high vacuum environment within the microscope column. However, the application of stain also requires dehydration of the sample, often leading to structural collapse. Further, uneven staining of the specimen limits the achievable resolution of the structure. Thus, alternative methods are required for high-resolution analysis of virus structures.

To preserve the natural structure of biological specimens in the electron microscope without the use of staining, Dubochet et al. (*26*) demonstrated successful cryogenic electron microscopy (cryo-EM) by embedding and recording images of adenovirus particles in a thin layer of vitreous ('glass-like'), i.e. non-crystalline, ice. This embedding was achieved by plunging a thin layer of specimen into liquid ethane, which immobilizes the sample and keeps the solvent in a glasslike, non-crystalline state (Fig. 2.1). The sample must remain below -137°C at all times to prevent irreversible crystallization of the vitreous ice (*27*). Cryo-EM samples are especially susceptible to radiation damage so images are recorded at minimal electron dosage, typically on the order of 6-10 electrons per $Å^2$. Particle images obtained by cryo-EM can then be

combined with image processing and three-dimensional (3D) reconstruction methods to produce high-resolution 3D structures.

The ability to visualize biological samples in a native or near-native state is the primary advantage that cryo-EM offers over other structure determination methods, but oftentimes it is also the most feasible method. Large macromolecular complexes are generally refractory to crystallization and NMR analysis. In cryo-EM, sample purity is not as essential as in X-ray crystallography and NMR methods because obvious impurities can be distinguished from the specimen of interest. Another advantage of cryo-EM over crystallographic and NMR methods is the limited amount of sample required (μg vs. mg quantities) for structure determination. These conveniences allow cryo-EM samples to be prepared from a simple series of ultracentrifugation and concentration procedures that can be completed within a relatively short period of time. For instance, I successfully purified APV (Chapter 3) and CW02 (Chapter 5) from infected cell cultures using a combination of low-speed centrifugation and CsCl-based ultracentrifugation. Following purification, the samples were dialyzed against native buffers to remove the CsCl and retain specimens in a native-like environment.

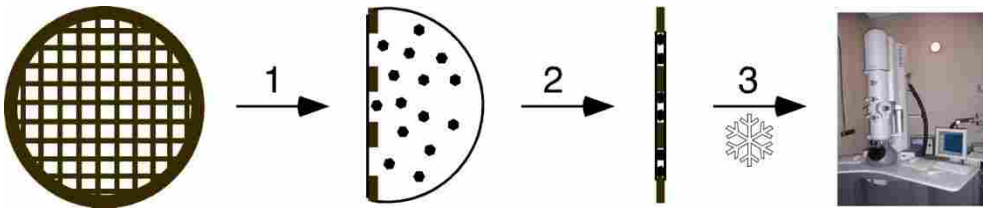


Figure 2.1. Cryo-EM schematic. (1) Approximately 3.5 μl of sample is applied to a copper mesh grid (left). (2) The specimen droplet is made thin by blotting to minimize ice thickness prior to vitrification. (3) The grid is plunged into the cryogen (liquid ethane) and can then be examined under the transmission electron microscope (right). The specimen is placed in a holder that keeps the specimen at cryogenic temperatures.

2.2 Freestanding-particle analysis and the projection theorem

Images acquired by cryo-EM have a poor signal-to-noise ratio due to low contrast and the low electron dosage used. This problem is offset by a technique known as "single particle analysis," or "freestanding-particle analysis," which improves the signal-to-noise ratio by averaging hundreds to thousands of homogeneous images. In principle, constant features of the particle images are reinforced while randomly distributed noise and other features are minimized. Consequently, images must be properly aligned translationally and rotationally for optimal averaging.

The two-dimensional (2D) images obtained by transmission electron microscopy represent 2D projections of the 3D object, i.e. the summation of densities along the view direction, as in medical X-rays. The projection, or central-slice, theorem for a 3D object states that the Fourier transform of a 2D projection is a central slice of its 3D Fourier transform normal to the view direction (Fig. 2.2). Thus, a 3D reconstruction can be obtained by combining 2D projections of the object from a wide range of view angles (orientations). Ideally, cryo-EM micrographs are images of particles that are randomly oriented in the vitrified solution so that the entire 3D Fourier space is covered.

The 3D reconstruction procedure relies on accurately determining the origin (e.g., center of mass) and orientation for each particle image. A common method of determining origins and orientation is by the use of projection-matching algorithms. In general, these algorithms rely on matching the experimental images with 2D projections of the 3D object. A database of regularly spaced projections is generated from the object. In principle, finer spacings between each projection will produce a more thorough database. Each experimental image is then assigned an orientation by matching the best model projection with each image. The origin of each particle

image is determined by translational alignment with the projection. The newly determined origins and orientations are then used to produce a more refined 3D structure, from which an updated database of projections is created. The cycle is repeated until the quality of the reconstruction (resolution) converges without further improvement.

Highly symmetric biological molecules, such as icosahedral viruses (see Chapter 1), are particularly amenable to cryo-EM methods because individual particles contain multiple equivalent orientations and because their orientations are more easily determined. For instance, icosahedral particles contain 60 redundant (asymmetric) units and thus each view has 59 other icosahedrally-related equivalents. Therefore, each icosahedral particle potentially consists of 60-fold more data than an asymmetric particle. The convenience of imposing icosahedral symmetry has produced several 3D reconstructions at near-atomic resolution (see (*28*) for a comprehensive review). Other factors that influence the quality of cryo-EM based 3D reconstructions include the number of particle images, defocus values of images, accuracy of orientation assignment to images, structural behavior of the imaged specimen (e.g., its rigidity or flexibility), and so forth.

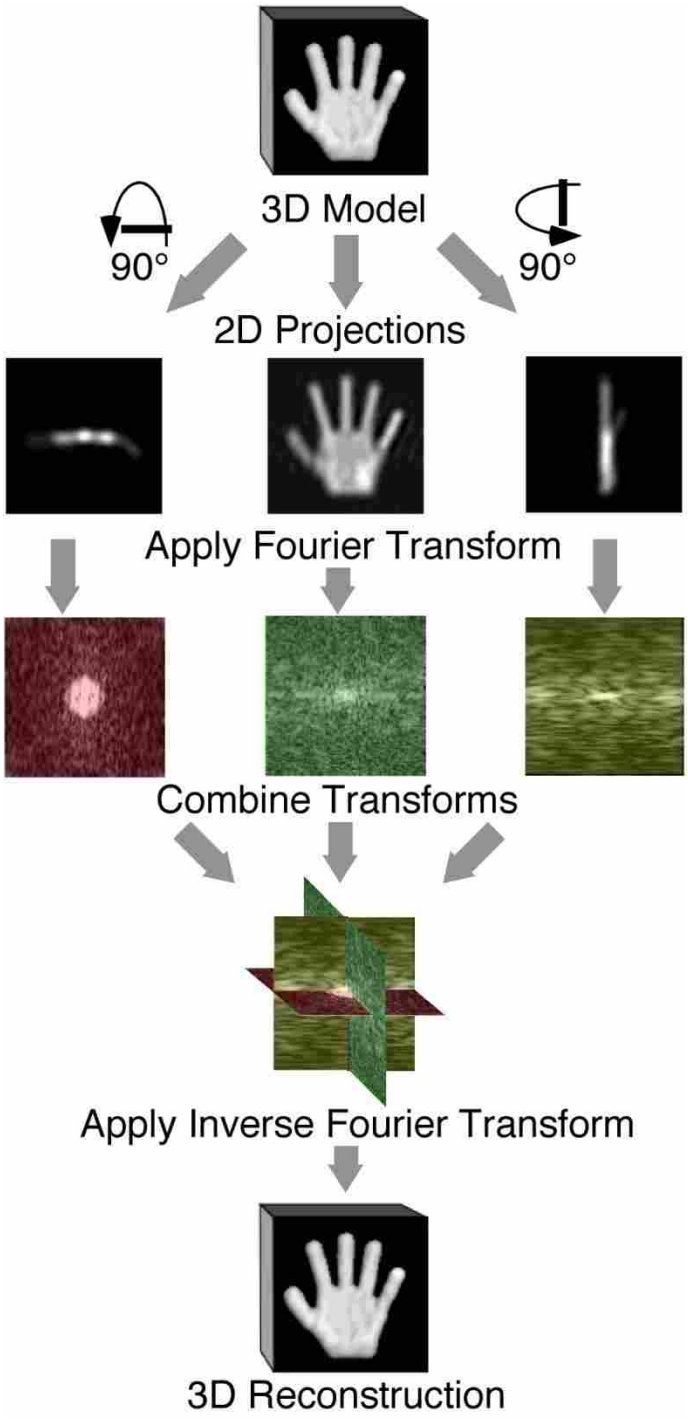


Figure 2.2. The projection, or central-slice, theorem. Each 2D image, or projection, represents a 'slice' that crosses the center, or origin, of the 3D Fourier transform. If the orientations of each slice are known, they can be assembled as a 3D Fourier transform. The 3D Fourier transform can be inverted to produce the 3D reconstruction. The ideal dataset will consist of projections covering the entire Fourier space and not just the orthogonal views shown here. Figure adapted from (*29*).

2.3 Challenges in cryo-EM methods

2.3.1 Model bias

As described in the previous section, successful cryo-EM based reconstructions generally rely on matching particle images to projections of the 3D object—but how are projection-matching algorithms initiated if the 3D structure is unknown? Most initial models are derived from indirect data, such as a 3D reconstruction of a related macromolecular complex, a map of a related X-ray or pseudoatomic structure, or a geometric construction. Yet, models derived from indirect data can potentially induce bias in the reconstruction process because they may have significant differences from the imaged particles. Further, the characterization of novel structures is limited by model-based approaches because models may not exist for the particle of interest. Thus, the ideal scenario is to generate an *ab initio* model directly from the imaged particles.

Recent demonstrations have shown the successful generation of low-resolution icosahedral (*30*) and asymmetric (*31*) structures without the need of a starting model. The structures were computed simply by assigning arbitrary orientations to a subset of particle images then iteratively refining the orientations. The correct structures often emerged after several refinement cycles despite starting with just a sphere-like blob. Thus, the structures derived from the 'random model' methods (RMMs) could be used as unbiased starting models for computing 3D reconstructions. I successfully used the RMM to generate starting models of polyomaviruses (Chapter 3). Further, applying the RMM to cryo-EM images of STMV resulted in a different structure than that previously solved by crystallography (Chapter 4). Finally, my characterization of halophage CW02 was dependent on the RMM because no prior model existed for the novel virus (Chapter 5).

2.3.2 Structural heterogeneity

Structural heterogeneity is among the most challenging limitations of high-resolution cryo-EM imaging. Traditional free-standing particle analysis relies on the averaging of identical, homogeneous structures that differ only in their orientation with respect to the electron beam. Yet, biological macromolecules are not static entities and may be inherently flexible and/or adopt different conformations. Thus, averaging heterogeneous populations will result in blurry, and perhaps uninterpretable, structures.

Despite the complications introduced by structural heterogeneity, one major advantage of cryo-EM is its ability to capture structural information of a specimen at different conformational states. By separating images of heterogeneous particles into homogeneous classes, one can compute several different 3D reconstructions from a single dataset of particle images. Nevertheless, the development of methods addressing particle heterogeneity remains one of the most important research topics in the cryo-EM field (*28*).

I encountered severe structural heterogeneity in my initial cryo-EM analysis of APV (see section 1.5.1 and Chapter 3). APV particles existed in an unexpectedly broad range of sizes and initial reconstructions were poorly resolved. Yet, the heterogeneity proved to be fortuitous as it revealed that APV lacks a conserved motif in its capsid structure compared to mammalian PyVs. I was then able to produce homogeneous APV particles via treatment with L-arginine, which yielded a sub-nanometer resolution reconstruction.

2.4 Modeling and fitting

Many detailed features of macromolecular complexes can be resolved at resolutions better than 5-Å, such as the turns and grooves of α-helices and strand separation in β-sheets. Yet for the most part, cryo-EM methods yield structures that are limited to resolving broad, tertiary structures of proteins or protein domains. Consequently, interpreting cryo-EM reconstructions is often challenging especially if the structures contain multiple domains.

The interpretation of low-resolution cryo-EM structures can be assisted by related high-resolution structures, typically derived by X-ray crystallography and NMR methods. Atomic-resolution models can be fitted into the electron density maps using map visualization software, such as UCSF Chimera (*32*). The fitting can be achieved manually ("by eye"), though manual efforts may be affected by user bias. Automated fitting provides a less biased result and the user can confirm the quality of the fit. The atomic structures can be fitted as a single entity (rigid-body fitting) by testing all possible positions and orientations (six-dimensional search, three translational and three rotational). A typical obstacle encountered during rigid-body fitting is a different conformation present between the EM map and the fitted atomic structure. Macromolecular complexes may adopt different conformations between cryo-EM and crystallization conditions. This has recently given rise to the development of flexible fitting methods (e.g., Molecular Dynamics Flexible Fitting (*33*)), which allows manipulation of local structures after an initial rigid-body fit.

An example of the power of fitting was shown in the 7.9-Å resolution cryo-EM structure of bacteriophage ϕ29. The structure appeared to consist of two domains: one with an HK97-like fold (see section 1.4.2) and the other with an immunoglobulin-like domain (Fig. 2.3) (*34*). Even

though the capsid protein sequences of ϕ29 and HK97 share < 20% identity, the authors used rigid-body fitting of the two separate domains to construct a single pseudo-atomic model.

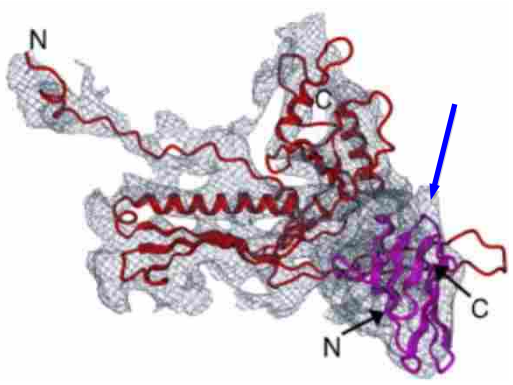


Figure 2.3. Pseudo-atomic model of bacteriophage ϕ29. The HK97-like (red) and immunoglobulin-like (purple) domains were fitted into the ϕ29 reconstruction. Note the clashing between the elongated HK97 loop and immunoglobulin-like domain (blue arrow), suggesting the fit in this region is incorrect. Figure adapted from (*34*).

The resolution of the reconstruction often determines the quality of the pseudo-atomic model obtained by fitting. Structures with better resolution have a lower degree of uncertainty about the positions of the atoms. Yet, caution should be taken because a good fit does not necessarily translate to a reliable structure; for instance, the combination of the HK97-like and immunoglobulin-like domains in the pseudo-atomic structure of ϕ29 capsids led to unrealistic clashing between coordinates (Fig. 2.3). Further, flexible fitting of proteins may cause steric torsions and other distortions that are improbable in nature. In any case, the user should avoid over-interpretation and "over-fitting" of lower resolution structures. Even for near-atomic resolution structures derived by cryo-EM methods, model building in the absence of existing comparable models is a challenging undertaking (*35*)—even the 3.88 Å cryo-EM reconstruction of cytoplasmic polyhedrosis virus only showed the main-chain Cα backbone of the capsid structure (*36*).

2.4.1 Homology modeling – a sequence-based approach

Sequence alignments of nucleic acids and proteins are common, powerful tools to infer evolutionary relationships among sequences. Because protein structure is more conserved than DNA sequence, detectable levels of sequence similarity may imply significant structural agreement (*37*). If the 3D structure of one of the protein sequences is known, then it can serve as a "template" by which an atomic-resolution "homology model" can be constructed for a related "target" protein. Homology modeling can be especially useful for providing a starting model of a macromolecular complex, which can then be verified by experimental procedures, including cryo-EM.

The quality of the homology model is dependent on the quality of the sequence alignment and template structure. As a rule of thumb, a reliable homology model can be constructed if the target and template share more than 50% sequence identity (*38*). The method simply requires a single amino acid sequence of the target protein as input data, after which the software will select suitable templates based on BLAST searching and automated sequence alignment. For sequences that share 20-35% sequence identity (also known as the "twilight zone" of sequence identity (*39*)), homology models can be constructed through the use of multiple sequence alignments. The use of several sequences allows the user to test several alternative alignments and evaluate the quality of the resulting models in order to achieve an optimal result. Sequences with less than 20% identity are not recommended for homology modeling analysis.

The pseudo-atomic resolution structures attained by homology modeling can then be verified by fitting into lower-resolution electron density maps. I constructed a homology model of the APV major capsid protein using the atomic coordinates of murine polyomavirus as template (see Chapter 3). The reliability of the model was reinforced by the high sequence

identity (41%) between the two proteins. Finally, the homology model was fitted into the sub-nanometer resolution reconstruction of APV, which suggested the model was accurate.

2.5 Other methods

Interpretation of cryo-EM structures can be aided by data obtained from other structural methods, such as modeling and fitting (described above), as well as from other non-structural methods, including genomic analysis and mass spectrometry. For instance, the 3D structure of CW02 (Chapter 5) was solved before its genome was sequenced. Thus, initial characterization of CW02 was limited to its structural features, e.g. *T* number, dimensions of the capsid, approximate number of protein subunits, and to a limited extent, structural relationship with other viruses. The sequencing of the CW02 genome enabled prediction of putative open reading frames. Each possible open reading frame was then used to determine homologous proteins among other viruses. The protein composition of mature CW02 particles was then confirmed by mass spectrometry, which matched peptide fragments to the CW02 database. My analysis revealed evolutionary relationships between CW02 and other phages isolated from diverse aquatic environments. This analysis also confirmed my initial speculation that the CW02 capsid has a conserved structural fold with other tailed, dsDNA bacteriophages.

Chapter 3

The structure of avian polyomavirus reveals variably sized capsids, non-conserved inter-capsomere interactions, and a possible location of the minor capsid protein VP4

3.1 Introduction

Polyomaviruses are widespread and infect mammalian and avian hosts. Virions are non-enveloped particles with a diameter of ~50 nm and contain a circular, double-stranded DNA (dsDNA) genome of ~5,000 base pairs. The structures of murine polyomavirus (MPyV) (*40-44*), simian virus 40 (SV40) (*44-48*), and human BK polyomavirus (BKV) (*49*) have been examined by means of X-ray crystallography or electron microscopy.

Mammalian polyomavirus capsids consist of 360 copies of the major capsid protein, VP1, that assemble into *T* = 7*dextro* (*T*, triangulation number) icosahedral shells comprised of sixty hexavalent and twelve pentavalent VP1 pentamers (*41, 44, 46*). Each pentamer contains five β-jellyroll folds formed by inter-digitating β-strands. Adjacent pentamers are interconnected via invading, C-terminal extensions, or 'arms' (*44, 46, 50*). The invading arms of five of the six quasi-equivalent VP1 monomers begin with an α-helix, which assemble into two- or three-helix "bundles" at hexavalent-hexavalent and pentavalent-hexavalent boundaries, respectively. In all VP1 subunits, the arm donates a β-strand to an adjacent capsomere. Calcium ions are necessary for capsid stabilization (*42, 44, 46*) and presumably support the invading arm (*44*).

Three of the six arms end in β-hairpins formed by strands K and L (*44, 46, 50*). Initially, these hairpins were proposed to plug holes in the capsid (*46*). More recent studies have shown, however, that partial truncation of the KL residues results in non-uniform particle sizes (*51, 52*), suggesting an assembly role.

Polyomavirus capsids also contain two minor structural proteins, VP2 and VP3. VP3 and the C-terminal two-thirds of VP2 have identical sequences. Crystallographic studies demonstrated that either VP2 or VP3 (VP2/3) inserts in a prong-like fashion into the axial cavities of the VP1 pentamers (*53, 54*). VP2/3 is highly disordered except for a conserved, 29 amino-acid stretch at its C-terminus (*54*). In this region, a conserved α-helix interacts with one of the five VP1 monomers near the base of the capsomere cavity (*53, 54*).

Mammalian polyomaviruses usually exhibit host- and tissue-specific tropism and are generally innocuous towards non-immuno-compromised hosts. Conversely, avian polyomavirus (APV) causes acute, sometimes fatal, disease in a variety of species and orders (*55-57*) and was detected in nearly all organs of infected birds (*20*). A phylogenetic analysis of polyomavirus genomes suggested that bird and mammalian polyomaviruses separated early in their evolutionary history and co-evolved with their hosts (*58, 59*). Differences between mammalian and bird polyomaviruses led to the proposal to separate them into different genera (*55, 60*). These distinctions include disparate pathology, differences of T-antigen binding to host genome, and the presence of a unique open reading frame in bird viruses (*23*). (For a comprehensive review comparing bird and mammalian polyomaviruses, see (*55*).)

The unique open reading frame in bird polyomaviruses encodes the minor capsid protein VP4 (176 amino acids), and its shorter, non-capsid-associated variant, VP4Δ (112 amino acids) (*61*). APV VP4 (hereafter, VP4) is a late gene product of APV (*61*) and has no sequence

similarity with SV40 VP4, which is a non-structural protein that triggers cell lysis and, like VP3, is derived from the VP2 open-reading frame (*62*). Expression of avian VP4 induces apoptosis in Sf9 insect cells and in chicken embryonic cells (*63*), though apoptosis is yet to be reported in infected birds. VP4Δ also induces apoptosis, though the protein is not incorporated into capsids (*63*). APV mutants that lack VP4 exhibit reduced virulence, suggesting that VP4 may contribute to acute pathogenesis (*23*).

VP4 likely has at least two roles in particle assembly. First, VP4 interacts with both VP1 and dsDNA, suggesting that VP4 participates in genome packaging (*61*); also, VP4 co-localizes with VP1 and VP3 in the nucleus of infected chicken embryonic cells (*64*). The VP4 sequence contains a region enriched with basic amino acids near a putative leucine-zipper (i.e. coiled-coil) motif, suggesting dimerization and DNA binding (*61*), cf. (*65*). Second, deletions in the VP4-encoding region of APV resulted in a high proportion of misassembled particles, suggesting VP4 facilitates capsid formation (*23*). However, the location and arrangement of VP4 within the capsid is unreported.

Here, we used cryogenic electron microscopy (cryo-EM) and other complementary techniques to investigate the three-dimensional (3D) structure of APV and compare it to the structures of SV40, SV40 virus-like particles (VLPs), and human JC polyomavirus (JCV). APV is very similar to the mammalian polyomaviruses except for its size variation, its lack of KL β-hairpin density, and the presence of a larger, electron-dense mass within the axial cavity of the VP1 pentamer, adjacent to the nucleohistone core. Our results suggest that the shorter, non-conserved C-terminus of APV VP1 affects capsid size and conformation and the KL β-hairpin in other polyomaviruses helps lock capsids in the stable conformation. The larger density associated with each APV capsomere may include VP4.

3.2 Materials and Methods

3.2.1 Cultivation and purification of APV

APV (strain, BFDV-5, GenBank accession number AF241170) was cultivated as previously described (*66*). It was purified according to the protocol used by Johne and Müller (*61*); briefly, infected chicken embryonic cells were sonicated on ice in the presence of Freon-TF. After removing cellular debris by low-speed centrifugation, the supernatant was loaded onto a two-step, CsCl density gradient (ρ = 1.3 and 1.4 g/ml). Two well-separated viral bands were observed. The top band (lowest density) had a higher percentage of empty particles than the bottom band (data not shown). Particles from the bottom band were used in all subsequent analyses. Viral bands were collected after spinning at 53,000 × g for 3 hours, and CsCl was removed by dialysis (D-tube Mini Dialyzers, MWCO (molecular weight cutoff) 12-14 kDa, Novagen, Gibbstown, New Jersey, USA) and the sample was retained in GP buffer (200 mM NaCl, 20 mM Tris-HCl, 1 mM $CaCl_2$ buffer, pH 7.4).

L-arginine has been used to reduce protein aggregation (*67*), and we treated purified APV with GP buffer plus 250 mM L-arginine (APV+R) by dialysis (D-tube Mini Dialyzers, MWCO 12-14 kDa, Novagen) to stabilize and de-aggregate particles. The pH of the solution was 10.7. The sample was subjected to centrifugal filtration (YM-100 centrifugal devices, Millipore, Billerica, Massachusetts, USA) for a final washing and concentrating step. Particle deaggregation also was attempted by treating samples with mild detergents, 0.1-0.5% Triton X-100 or Igepal (Sigma-Aldrich, St. Louis, Missouri, USA), and adjusting the pH between 4.0 and 10.7.

Purified APV was subjected to SDS-PAGE and immunoblotting as previously described (*60, 61*). Rabbit polyclonal antisera αAPV (specific for APV structural proteins VP1, VP2, VP3 and VP4) and αVP4 (specific for VP4) were used for immunodetection of the proteins (*61*).

3.2.2 Cultivation and purification of SV40 and SV40 VLPs

SV40 virions were cultivated and purified as described previously (*45*). SV40 VLPs were produced and isolated from Sf-9 cells using a recombinant baculovirus system as reported (*48*). Briefly, wild-type VP1, and recombinant VP2 and VP3 fused with enhanced green fluorescent protein were cloned individually into pFastBac-1 plasmid vectors (Invitrogen, Carlsbad, California, USA) and introduced into BacMid vectors (Invitrogen). The BacMid vectors were transfected into Sf-9 cells to produce recombinant baculovirus. Sf-9 cells were then infected with plaque-purified, recombinant baculoviruses to generate recombinant SV40 VLPs. The insect cells were harvested three days after infection and cell lysates were prepared via sonication. The resulting VLPs were purified on a 20-50% CsCl density gradient and the virus-containing band was further concentrated with a 37% CsCl density gradient. Finally, the fractions containing VLPs were collected and dialyzed against a buffer containing 20 mM Tris (pH = 7.9) and 50 mM NaCl.

3.2.3 Cultivation and purification of JCV

The Mad-1SVEΔ strain of JCV was used in this study (*68*). JCV was propagated in SVG-A cells as previously described (*69*). JCV was purified using a modified version of previously published results (*70-72*). At 16-19 days post infection, cells were scraped from the flasks in the presence of growth media, and then pelleted by centrifugation. The virus-containing pellet was resuspended in 10 mL of Buffer A (10mM Tris, 50mM NaCl, 0.1 mM $CaCl_2$, pH 7.4) supplemented with a protease inhibitor cocktail (Sigma-Aldrich, St. Louis, Missouri, USA). The

pellet was frozen and thawed three times, after which the lysate was treated with 0.5M HEPES buffer (pH 5.4) to lower the pH to 6.0. Next, five units of type V neuraminidase (Sigma-Aldrich) were added to the lysate for 1 hr at 37°C, then the pH was raised back to 7.5 with 0.5M HEPES (pH 8.0). $CaCl_2$ was added to a final concentration of 0.5mM and cell lysates were treated with 1000 units of DNase for 1 hr at 37°C (New England Biolabs, Ipswitch Massachusetts, USA), after which the lysate was sonicated. The suspension was then pelleted at 4°C, resuspended in 5 mL of Buffer A, sonicated, and centrifuged again as above. The supernatants were pooled following a final round of resuspension and sonication. The virus-containing supernatant was pelleted through a 20% sucrose cushion in at ~155,000 × g for 2.5 hrs (Beckman-Coulter, Brea, California, USA). The resulting pellets were pooled together and resuspended in Buffer A containing CsCl to a final density of 1.34 g/ml; this mixture was centrifuged at 152,000 × g for 16 hr. The virus-containing band was extracted and then dialyzed against either Buffer A alone or Buffer A with 250 mM L-arginine (JCV+R, pH 10.7) (Slide-A-Lyzer, 10 kDa MWCO, Thermo Fisher Scientific, Rockford, Illinois, USA).

Analysis of minor capsid protein retention in polyomaviruses following arginine treatment to assess whether VP2 remained associated with the virion after arginine treatment, JCV+R was purified by size-exclusion chromatography (SEC) on a Sephacryl S100 column (GE Healthcare, Piscataway, New Jersey, USA) and absorbance (λ = 280 nm) of each fraction was measured. Fractions corresponding to intact virions were pooled and loaded onto 10% SDS-PAGE gels. Gels were color silver stained (Pierce), and imaged on a Chemidoc XRS gel scanner (Bio-Rad, Hercules, California, USA). The presence of VP2 was compared to a non-SEC-purified control to determine whether VP2 was dissociated from virions after L-Arg treatment.

APV, APV+R, SV40 VLPs, and SV40 VLPs+R were dissociated into their pentameric subunits (*73*) by using overnight dialysis in a buffer containing 200 mM NaCl, 5% glycerol (v/v), 10 mM dithiothreitol, 10 mM EDTA, and Tris-HCl, at pH 7.4 and 4 ºC ("+R" buffers also contained 250 mM L-arginine). The samples were then placed on top of a 5-40 %, 5-30 %, or 5-20 % sucrose gradient made by gently layering sucrose solutions in equal amounts and allowing the layered solution to equilibrate overnight at 4 ºC. The samples were centrifuged at ~285,000 × g for 3 hours at 4 ºC and equal volume fractions were extracted from the top of the gradient. Fractions were analyzed via SDS-PAGE.

3.2.4 Cryo-EM and image reconstruction

SV40 virions were imaged as described previously (*45*, *47*). For all other samples, 3.5 µl of purified virus solution was placed on a glow-discharged, holey-carbon-coated copper grid, blotted, and plunge frozen in liquid ethane with an FEI Vitrobot (Hillsboro, Oregon, USA). Specimens were transferred to a Gatan 626 cryoholder (Pleasanton, California, USA) cooled with liquid nitrogen. Cryo-EM images were recorded at either 200, 250, or 300keV and at 39,000× magnification via low-dose methods at objective lens underfocus levels ranging from 0.5–4.0 µm on Kodak SO-163 film in an FEI Tecnai F30 transmission electron microscope (Hillsboro, Oregon, USA) (Table 3.1).

All electron micrographs were digitized on a Nikon Super Coolscan 9000 ED scanner. Particle images were extracted from electron micrographs using X3DPREPROCESS (*74*) (Table 3.1). Using Bsoft (*bshow* and *bctf* functions) (*75*), we determined contrast transfer function signal and decay parameters and applied them to correct the images. Origins and orientations of the extracted particles were determined using the model-based technique of *PFT2* (*76*), which was adapted to use phase and amplitude information in orientation selection (*31*). For APV and

SV40, a reference model for *PFT2* analysis was generated by the random-model method with imposed icosahedral symmetry (*30*). For JCV and SV40 VLPs, reference models were the same as the APV starting model. The final 3D reconstructions were calculated using *EM3DR2* (*77, 78*). Bsoft (*bresolve* and *blocres* functions) (*75*) was used to estimate the resolution of reconstructions (Table 3.1). Two reconstructions were computed, each from alternating particle images in the data set, and used to assess resolution. Quasi-local-resolution was determined by comparing 32^3 voxel (1.42×10^5 Å^3/pixel^3) sub-volumes (from the two reconstructions) centered at each voxel in the capsid region of the APV+R structure via a routine implemented in Bsoft (G. Cardone, unpublished) (*75, 79, 80*). Contour levels are given in terms of σ, which was calculated as the number of standard deviations relative to the average map density.

We calibrated the image pixel size by comparing radially averaged density plots (*81*) with Bsoft (*bradial* function) (*75*) (Fig. S2). Density computed from the atomic coordinates of SV40 (PDB ID, 1SVA, (*44*)) was our standard. Relative size factors were determined via a routine within *PFT2* that changes the size of the reference projection and determines the best correlation with respect to each particle image. Following size calibration, Bsoft (*bshow* function) (*75*) was used to estimate the lengths and volumes of densities found inside capsomeres.

3.2.5 Sequence and homology model analysis

Polyomavirus VP1 and VP2 sequences, obtained from GenBank (see Appendix for accession numbers), were aligned using ClustalW2 (*82*). We created APV VP1 homology models of each quasi-equivalent subunit (α, α', α'', β, β', and γ) using the SWISS-MODEL server (http://swissmodel.expasy.org) under the program's alignment mode (*38*). Bond energies were

minimized using UCSF Chimera (*32*), after which PROCHECK was used to check structural soundness by a Ramachandran plot (*83*).

The APV homology model was fit as a rigid body into the electron density map (*32*, *84*). Molecular images were produced using the UCSF Chimera package (*32*). Full capsid coordinates of murine polyomavirus (*50*), SV40 (*44*), and the APV homology model were assembled by applying icosahedral symmetry using UCSF Chimera (*32*) or the ViperDB oligomer generator (*85*). Hydrogen-bonding sites were determined using UCSF Chimera (*32*) with relaxed constraints of 0.4 Å and 20º, based on geometric criteria previously reported (*86*).

3.3 Results

3.3.1 Cryo-EM of polyomaviruses

We recorded cryo-EM images and computed three-dimensional reconstructions of several polyomavirus samples: APV, APV treated with 250 mM L-arginine (APV+R), SV40, SV40 VLPs, JCV, and JCV treated with 250 mM L-arginine (JCV+R) (Table 3.1).

Table 3.1. Microscopy and image reconstruction data.

Sample	Voltage (kV)[a]	Number of images	Reconstruction resolution[b] (Å)	Particle diameters (nm)	Plug length (Å)	Plug volume (kÅ^3)	EM Data Bank ID
APV 'tiny' (pH 7.4)[c]	250, 300	168[d]	-	< 46	-	-	-
APV 'small' (pH 7.4)	250, 300	434[d]	31	46-51	n/a[g]	n/a[g]	-
APV 'medium' (pH 7.4)	250, 300	570[d]	35	51-57	n/a[g]	n/a[g]	-
APV 'large' (pH 7.4)[c]	250, 300	56[d]	-	> 57	-	-	-
APV+R (0.25M arginine, pH 10.7)	200	5,338	11[f]	47-53	43±2	33±3	5180
APV_{empty}+R (empty particles)	200	671	18	47-53	39±3	23±3	5181
SV40 (pH 7.4)	80[e]	1,155	23	47-53	29±1	13±2	5187
SV40 VLPs (pH 7.4)	200	13,857	14	47-53	31±2	18±4	5182
JCV (pH 7.4)	200	1,139	19	-	32±1	17±2	5183
JCV+R (0.25M arginine, pH 10.7)	200	1,044	18	-	n/a[g]	n/a[g]	5184

[a] Except where noted, micrographs were recorded on a FEI Tecnai F30 microscope at a nominal magnification of 39,000x.

[b] Resolution was estimated by using the Fourier shell correlation criterion (*87*) cutoff of 0.5.

[c] Datasets failed to produce consistent reconstructions.

[d] The relatively low number of particles used in the APV reconstructions (without arginine) was a result of difficulties in extracting individual particle images from particle aggregates (particles must be freestanding). Reconstructions of APV 'small' and 'medium' are not shown.

[e] SV40 data set recorded on a Philips EM420 electron microscope at a nominal magnification of 36,000×.

[f] By means of a quasi-"local-resolution" test (G. Cardone, in preparation), we determined that the core, β-jellyroll fold was rendered at ~8.5–10-Å resolution (data not shown). Similar resolution was achieved in the αC helices that stabilize intercapsomere interactions. Other more flexible regions, such as those corresponding to exterior loops, were less well resolved (~10–14-Å). In principle, the more rigid a structure is, the higher the resolution that can be attained in crystal structure and cryo-EM studies.

[g] Plugs were not observed in the JCV+R reconstruction, but plugs were observed in the non-arginine treated APV 'small' and 'medium' reconstructions. The poor resolution and size uncertainties made measurement of the 'small' and 'medium' APV plugs unfeasible.

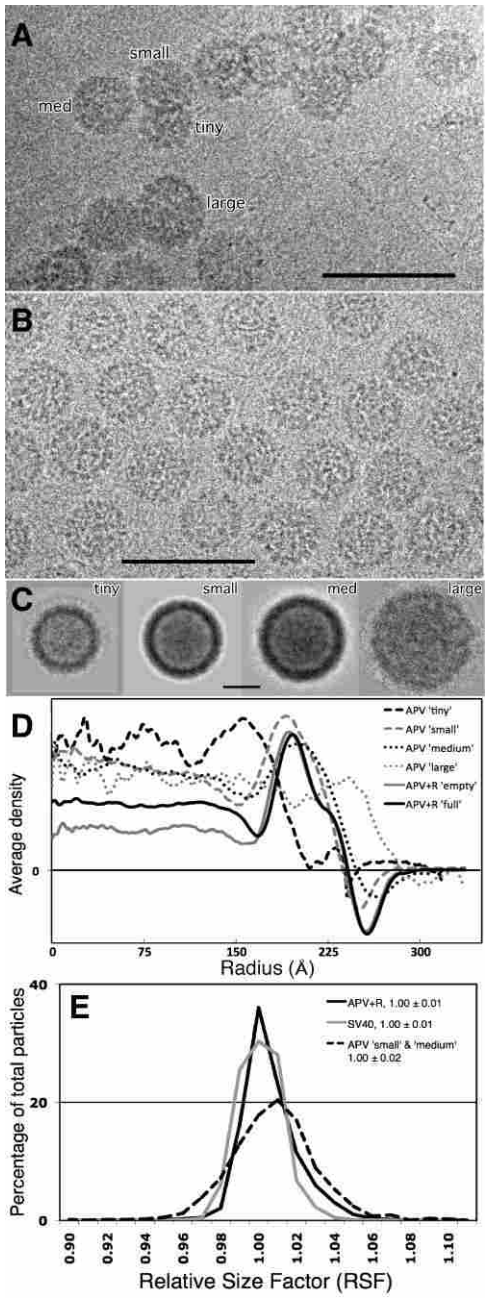


Figure 3.1. Effect of L-arginine on APV. (A) Cryo-EM image of APV shows particle aggregation and size variation (particles classified according to size). (B) Cryo-EM image of APV after arginine treatment shows significantly less particle aggregation and size variation than those without treatment (A). Scale bars, 100 nm (for A and B). (C) Image averages of four classes of variably sized APV particles observed in preparations like that shown in (A). Particle images were aligned, summed, and normalized via a procedure described previously (*81*). Scale bar (for C), 20 nm. (D) Radial density profiles of averaged images from C and from images of arginine-treated APV. (E) Relative size factor (RSF) distributions as assessed by *PFT2* of three different cryo-EM data sets, with 1.00 set as the standard based on the MPyV reference model. Legend shows sample type and the mean ± standard deviation for each data set (bin size = 0.01).

3.3.2 Cryo-EM of APV reveals variably sized capsids

Cryo-EM analysis of APV revealed extensive particle aggregation and variable particle sizes (Fig. 3.1A). With murine polyomavirus (MPyV, ~51 nm diameter) (*50*) serving as a control, we determined that relative size factors (RSFs) of APV particle images ranged between 0.75−1.27 (diameter range 38–65 nm). Unaggregated APV particles with circular profiles were divided into four size classes, designated 'tiny', 'small', 'medium', and 'large' (Table 3.1, Fig. 3.1C). Conversely, SV40 particles exhibited a narrower size distribution (Fig. 3.1E). Each particle class appeared to have filled cores (Fig. 3.1D), suggesting that each contained DNA. We attempted to compute a 3D reconstruction from the four classes of APV particle images, but only the 'small' and 'medium' classes yielded consistent results—similar $T = 7$, icosahedral structures at 31- and 35-Å resolution, respectively (Table 3.1).

3.3.3 Cryo-EM of arginine-treated APV (APV+R)

We attempted to improve the resolution in our APV reconstructions by minimizing aggregation of APV particles. Aggregation persisted over a large pH range (4−11), even in the presence of mild detergents. Higher resolution reconstructions were not obtained if the 'small' and 'medium' classes of particle images were pooled into a single dataset or if a narrower range of particle sizes (0.99−1.01 RSF) was used (data not shown).

Treatment of APV with 250 mM L-arginine (APV+R) markedly reduced particle aggregation and substantially narrowed the distribution of APV particle sizes to a range that is comparable to that observed for SV40 particles under mild buffer conditions (Fig. 3.1B,E). In addition, the average size of the APV+R particles lies between the 'small' and 'medium' classes on non-arginine treated APV (Fig. 3.1D). The addition of arginine led to a cryo-reconstruction of APV+R at ~1-nm resolution (Table 3.1, Fig. 3.2). Addition of L-arginine raised the pH of the

APV sample from 7.4 to 10.7. Readjustment of the pH to 7.4 led to reaggregation of the particles (Fig. S3.1). Furthermore, aggregation was not reduced by raising the pH of the APV solution to 10.7 in the absence of arginine (Fig. S3.1). Although a change from pH 7.4 to 10.7 might cause significant conformational changes in some proteins, a reconstruction of JCV treated with 250 mM L-arginine (JCV+R) showed a similar VP1 structure as untreated JCV (Fig. 3.2). Arginine treatment does appear to lower the density inside the particle (Figs. 3.1D, S2).

The APV+R reconstruction closely resembled cryo-reconstructions of SV40 and JCV (Fig. 3.2, 3.3C). The APV+R capsid has the same quaternary structure as mammalian polyomaviruses: pentameric, barrel-shaped capsomeres arranged with T=7 icosahedral symmetry. Inter-capsomere densities are high where αC helices from two (β-β') or three (α, α', and α") VP1s are known to interact and where twofold-related hexavalent capsomeres lie close together (*44, 46, 50*).

3.3.4 Sequence analysis and homology modeling of APV VP1

We investigated the basis of APV size variation by comparing the VP1 sequences of several polyomaviruses. With the exception of the more distantly related WU and KI polyomaviruses, alignment of eighteen polyomavirus VP1 sequences revealed 41−62% amino acid identity between bird and mammalian sequences (Fig. S3.3, S3.4). High sequence conservation occurs in the β-jellyroll fold, the disulfide bond-forming residues, and most motifs that tether adjacent capsomeres (Fig. S3.3).

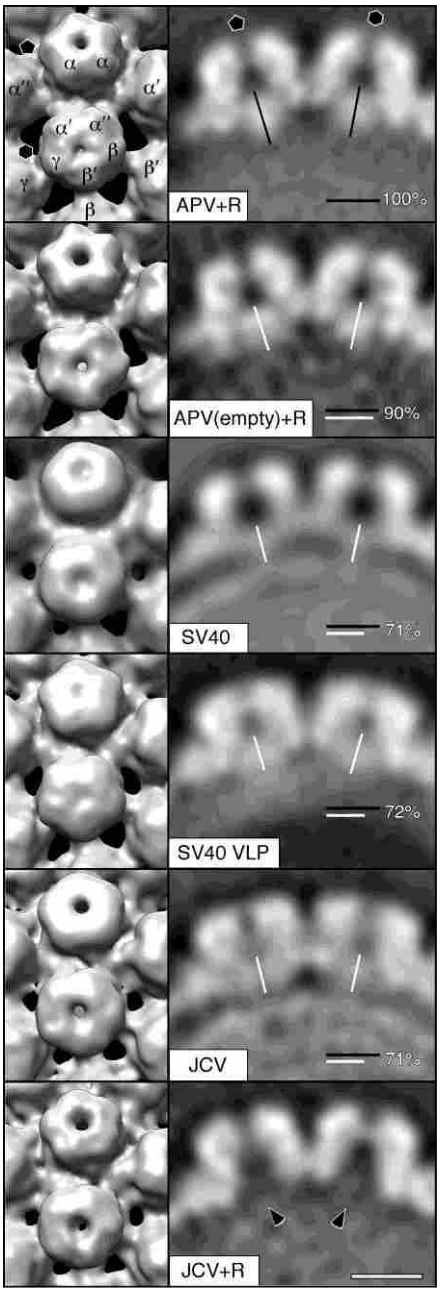


Figure 3.2. Views of pentavalent and hexavalent capsomeres based on cryo-EM reconstructions of APV+R, APV+R empty capsids, SV40, SV40 VLPs, JCV, and JCV+R all rendered at 25-Å resolution to allow for direct comparison. Left column, surface rendering views of fivefold symmetric pentavalent (top, filled pentagon) and non-symmetric hexavalent (bottom, empty hexagon) capsomeres and inter-capsomere connections rendered at 1σ density level. Top, VP1 monomers labeled according to Liddington et al. (*46*). Right column, cross-sectional density slices of pentavalent (left) and hexavalent (right) capsomeres. Plug densities were observed at the VP1 pentamer-dsDNA interface, but not when JCV was treated with arginine (black arrowheads). Black (APV+R) or white lines overlaid against plug densities show the measured length of the plug-like density. Lines at the bottom right corner of each panel compare the average pentavalent and hexavalent plug length against the APV+R plug length, with percentage given. Scale bar (striped), 5 nm, for right column only.

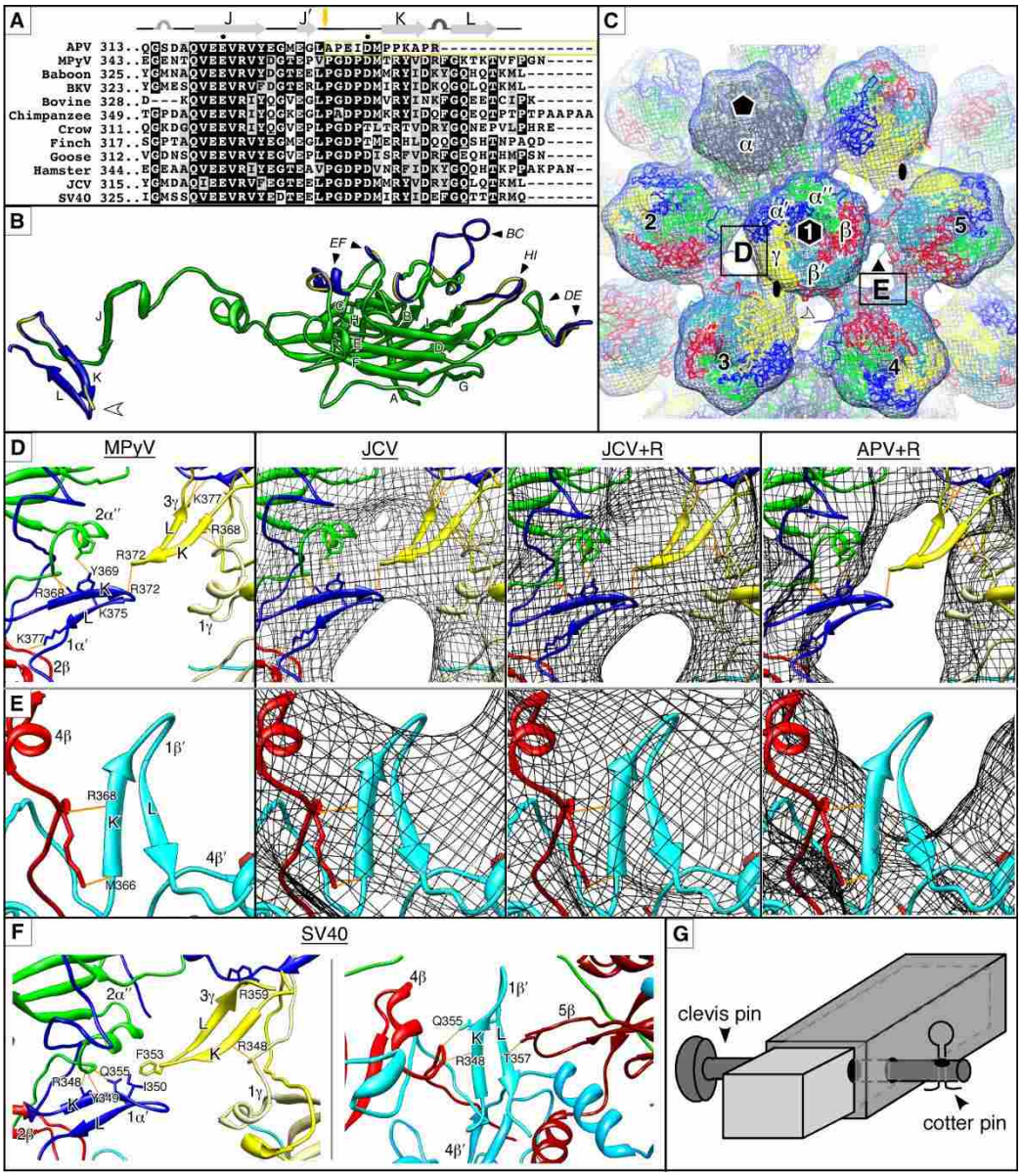


Figure 3.3. The structure of APV VP1 compared to other polyomaviruses. Where colored, VP1 subunits are α (gray), α' (dark blue), α'' (green), β (red), β' (turquoise), and γ (yellow).

(A) C-terminal alignment of 12 variants of polyomavirus VP1. Secondary structure, based on MPyV coordinates (*50*), is shown above alignment. Black boxes show identical residues, where at least 50% of the aligned amino acids are conserved. Gray boxes show amino acids with conserved properties. Dots show putative Ca^{2+} binding residues. The APV sequence is not conserved beyond the J´ strand (green arrow) and terminates at the loop between strands K and L. Virus name or host is given on left.

(B) Ribbon diagrams of the homology model of APV VP1 γ subunit (yellow) superimposed over the crystal coordinates of the MPyV VP1 γ subunit (blue). Conserved structures are colored green. β-strands and loops labeled in non-italicized and italicized fonts, respectively. The BC, DE, EF, and HI loops lie on the exterior of the capsid. As shown in (A), the APV sequence terminates after the K β-strand. The C-terminus of APV VP1 (open arrowhead) does not end in a freestanding β-strand and likely does not follow the K β-strand of MPyV.

(C) APV homology model (wires) fit into 3D reconstruction (gray mesh) of APV+R. Representative α, α', α'', β, β', and γ subunits are labeled. A fivefold symmetric pentavalent capsomere is labeled with a filled pentagon. Non-symmetric hexavalent capsomeres (unfilled hexagon) contain multi-colored subunits and are numbered. Three- and twofold symmetry axes are labeled with triangle or oval, respectively. Boxes D and E and numbered capsomeres (1–5) refer to panels below. APV reconstruction was rendered at 1σ. C-terminal residues corresponding to the K β-

strand in MPyV and SV40 (α', β', and γ subunits, e.g. open arrowhead) protrude out of the APV+R reconstruction density, showing that the homology model in this region is incorrect.

(D,E) Inter-capsomere interactions of KL β-hairpins adjacent to twofold (D) and threefold (E) axes rendered at 1.2σ density level. MPyV panels show only the crystal structure (*50*). JCV and JCV+R panels show MPyV structure fitted into cryo-EM density maps and have density corresponding to the KL β-hairpin. APV+R panels show absence of density where the KL β-hairpin is located. Identified hydrogen bonds are orange with the KL residue labeled (see Table 3.2). Individual protein subunits are identified by the pentamer where the subunit originates (Arabic number) and by the subunit within the pentamer (Greek letter). Subunits with the same Greek letter are given a different shade of the same color.

(F) Inter-capsomere interactions seen in the SV40 crystal structure (*44*) shown as in (D), left, and (E), right. Some interactions are similar to those in MPyV but others differ (cf. Table 3.2).

(G) The KL β-hairpin is analogous to a mechanical cotter or retaining pin, which is a small pin that inserts into one or both ends of a larger and stronger clevis pin used to connect two objects. A cotter pin prevents removal of the clevis pin. In polyomaviruses, the KL β-hairpin functions like a cotter pin in the sense that it more securely fastens the stronger, remaining parts of the C-terminal arm (e.g. the J strand, αC, and Ca^{2+} chelation) that form intercapsomere connections. Therefore, interactions of the KL β-hairpin may serve as a final "lock" that secures pentamers in the stable, mature conformation.

The C-terminal, 12-residue sequence following the J'-strand (amino acids 332–343) is not conserved in APV VP1 and is truncated (Figs. 3A and S3). Other polyomavirus VP1 sequences are significantly longer and are generally conserved in this region. The shorter C-terminus of APV VP1 lacks residues that would normally form a terminal β-hairpin loop composed of strands K and L (KL β-hairpin). The non-conserved sequence is present among all three currently sequenced APV strains (data not shown), indicating that the "premature" truncation is not limited to a single isolate. If the APV VP1 stop codon were ignored or various frame shifts were introduced, subsequent codons did not code for amino-acid residues that were similar to those in the β-hairpin of other polyomavirus VP1s.

Other significant differences between the bird and mammalian VP1 sequences occur at the N-terminus and in four external loops (BC, DE, EF, and HI) (Fig. S3.3). The N-termini in mammalian polyomaviruses contain the nuclear localization signal and the DNA-binding domains, but these are not conserved in APV (*64*). Regions corresponding to external loops BC and HI contain residues that mediate receptor binding in SV40 (*88*) and MPyV (*89*).

High conservation of the core jellyroll fold and inter-capsomere connections provided a rationale for using the MPyV structure (*50*) as a template to create a homology model for

residues 21−343 of APV VP1 (Fig. 3.3B). The APV homology model fit nicely into the electron density map of the APV+R reconstruction, confirming structural conservation of the major jellyroll fold and *T* = 7*d*, pentameric quaternary structure (Fig. 3.3C). Inter-capsomere stabilization motifs, mediated by the C-terminal arm, are also structurally homologous. Differences between the APV and MPyV structures were seen in the external loops and KL β-hairpin (Fig. 3.3B). C-terminal residues following the J'-strand protrude out of the APV+R reconstruction, demonstrating that the homology model is incorrect in this region (Fig. 3.3C).

3.3.5 KL-hairpin interactions in mammalian polyomaviruses

Densities attributable to the KL β-hairpin interactions, as observed in the crystal structures of SV40 (*44*, *46*) and MPyV (*50*), were observed in all the moderate-resolution, 3D reconstructions of JCV and SV40 samples, but were absent in the higher-resolution APV+R 3D map (Figs. 2; 3D,E). The lack of hairpin density was likely not an effect of arginine treatment because the densities remained in the JCV+R reconstruction (Fig. 3.3D,E). Although the KL β-hairpin structures were observed in α', β', and γ subunits of the crystal structures of SV40 (*44*, *46*) and MPyV (*50*), some details of their structures have not been described. The KL-hairpins in MPyV provide additional stability to virions through several inter-capsomere interactions, primarily involving hydrogen bonds (Table 3.2; Fig. 3.3D,E). The KL-hairpins of SV40 involve similar inter-capsomere interactions (Table 3.2, Fig. 3.3F)—for instance, a conserved arginine in MPyV (R368) and SV40 (R348) hydrogen-bonds with a conserved lysine residue. The SV40 hairpins also participate in non-conserved interactions, such as the hydrophobic interface formed between I350 of a α' hairpin and F353 of a γ hairpin (Table 3.2, Fig. 3.3F). The stabilizing function of the KL-hairpin is non-existent in APV.

3.3.6 Axial density at base of VP1 pentamer is larger in APV+R than in other polyomaviruses

The axial cavity of each VP1 pentamer in the APV+R reconstruction is “plugged” by an electron-dense mass at the inner capsid surface (Fig. 3.2, right). Previous studies demonstrated that VP2/3 insert into the cavity in a prong-like fashion (*53*, *54*). Sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblot analyses of the APV+R sample used in our cryo-EM studies confirmed the presence of VP2 and VP3 in the virion (Fig. 3.4A).

The 3D reconstructions of SV40, SV40 VLPs, and JCV revealed similar, yet smaller, 'plugs' at the same location as in APV+R (Fig. 3.2, right). SDS-PAGE analysis also confirmed the presence of VP2 and VP3 in the JCV sample (Fig. 3.4B). The plug densities in all reconstructions spanned between the axial cavity of the VP1 pentamer and the nucleohistone core of the virion and were observed in both hexavalent and pentavalent capsomeres. The magnitude of these densities was similar to that seen in the VP1 portion of the capsomeres, indicating that most, if not all, of the 72 capsomeres in the capsid are plugged.

For each reconstruction, the approximate volume and length of the plugs in the hexavalent and pentavalent capsomeres were measured from density slices (Table 3.1; Fig. 3.2, right). The APV+R plugs were larger compared to those in the mammalian polyomavirus reconstructions, even though the polypeptide-chain lengths of VP2 and VP3 are nearly identical and the VP1-binding helix is conserved in all polyomaviruses (Fig. S3.5) (*54*). We computed a reconstruction of empty capsids of APV+R (APV_{empty}+R), and these also had larger plugs compared to SV40, SV40 VLPs, and JCV (Table 3.1, Fig. 3.2, right). On average, the plug occupied a volume of 28 (± 4) $\times 10^3$ $Å^3$ in the APV+R and APV_{empty}+R reconstructions. An

unsymmetrized model of APV+R was also computed (*31*), and this confirmed the presence of plugs in the same locations (data not shown). The plugs in SV40, SV40 VLPs, and JCV occupied significantly less volume ($16 (\pm 4) \times 10^3$ Å^3). Length measurements showed that APV+R and APV_{empty}+R plugs were 9 (± 4) Å longer than those in SV40 and JCV (Table 3.1, Fig. 3.2, right).

3.3.7 Effect of L-arginine on interactions between major and minor capsid proteins

Although the JCV and JCV+R capsid structures are nearly identical, our JCV+R reconstruction surprisingly showed no detectable densities in the axial cavities of the VP1 pentamers (Fig. 3.2, right). To see whether arginine treatment caused VP2/3 to be discharged from the JCV capsid, we used size-exclusion chromatography on a JCV+R sample, which should separate virions from any VP2 free in solution. No protein absorbance peaks (λ = 280nm) were detected in fractions following virion elution, indicating VP2 and VP3 were retained within the capsid (Fig. 3.4C). SDS-PAGE analysis of the virion fraction confirmed that VP2 and VP3 remained within the particles in a similar ratio as untreated JCV (Fig. 3.4B). Thus, while arginine treatment likely dissociates VP2 and VP3 from VP1, the minor capsid proteins are not ejected from the capsid. The increased pH caused by arginine treatment may be the basis for dissociation of VP2 and VP3; indeed, treatment of SV40 particles with DTT at pH 9.8 dislodged most VP2 and VP3 from VP1 (*90*).

3.3.8 Disruption of VP1 pentamer by EDTA, DTT, and L-arginine

For our SV40 VLP and APV samples, we investigated if minor capsid proteins remained connected to the VP1 pentamer after arginine treatment. Virions and VLPs were dissociated into VP1 pentamers using the required EDTA and DTT (*73*) with and without arginine. Solutions without arginine clearly showed co-sedimentation of VP1 and minor capsid proteins (Fig. S3.6). Surprisingly, solutions with arginine showed co-sedimentation of capsid proteins at a much

slower rate, indicating that the VP1 pentamer had been disrupted by treatment with EDTA, DTT, and L-arginine (Fig. S3.6A,B). Negative-stain electron microscopy of the dissociated APV capsids showed pentamers in the EDTA-DTT solution, but pentamers were not observed in the EDTA-DTT-arginine solution (Fig. S3.6C,D). Therefore, we were unable to assess the composition of minor capsid proteins that remain attached to VP1 pentamers after arginine treatment.

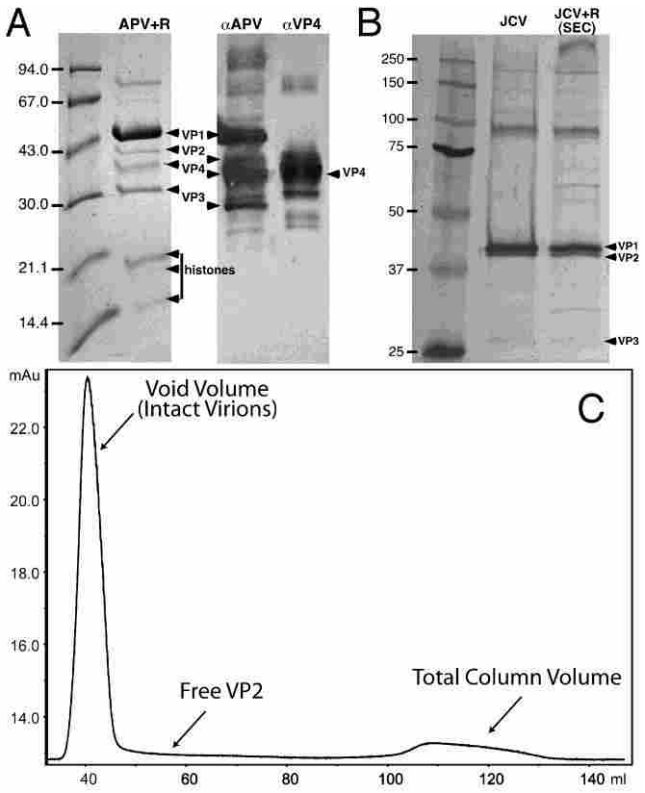


Figure 3.4. Minor capsid proteins after treatment with L-arginine: APV+R (A) and JCV and JCV+R (B,C). (A) Coomassie-stained SDS-PAGE (left) and immunoblot (right) analysis confirms the presence of VP4 in the sample used to image APV. In the immunoblot, antibodies to all APV structural proteins (left, αAPV) or VP4-specific antibodies (right, αVP4) were used. Note that our APV samples retained protein components, including histones, suggesting that the KL β-hairpin is not a plug as previously suggested (*46*). (B) Silver-stained SDS-PAGE gel confirms the presence of VP2 and VP3 in JCV virions without and with L-arg treatment. Second lane, JCV without arginine (VP2:VP1 ratio, 0.47). Third lane, JCV+R purified from S100 column (VP2:VP1 ratio, 0.48). High molecular weight bands are seen in both preparations and are likely cross-linked VP1 that did not dissociate during treatment with denaturing buffer. (C) Chromatography of JCV+R purified by size exclusion chromatography (SEC) on a Sephacryl S100 HR column. Plot shows absorbance (milli-absorbance units, mAu, at 280 nm) vs. accumulated fraction volume (ml). Intact virions are eluted in the void volume, whereas any free VP2 should have eluted later (arrow). Although we cannot exclude the fact that some VP2 could be lost on the column, the ratio of VP1-to-VP2 in the control and SEC lanes are nearly identical (see B), indicating that most VP2 remained with the virion after arginine treatment.

Table 3.2. Inter-subunit, inter-pentamer interactions that include a residue in K β-strand, hairpin turn, or L β-strand.[*,e]

	KL β-hairpin			Interacting Entity		
Virus	Subunit[a]	Residue	Group or Atom	Subunit[a]	Residue(s)	Group or Atom
SV40[b]	3γ	Phe353	hydrophobic[c]	1α'	Ile350	hydrophobic[c]
	3γ	Arg348	main-chain O	1γ	Lys29	main-chain NH
	3γ	Thr358	main-chain O	1α'	Tyr299	side-chain OH
	1α'	Gln355	side-chain NH	2α"	Leu26	main-chain O
	1α'	Arg348	main-chain O	2α"	Lys29	main-chain NH
	1α'	Arg348	side-chain NH	2α"	Leu26	main-chain O
	1β'	Arg348	main-chain O	4β	Lys29	main-chain NH
	1β'	Gln355	side-chain NH	4β	Val27	main-chain O
	1β'	Thr357	main-chain NH	5β	Asp101	side-chain O
MPyV[d]	3γ	Arg372	main-chain O	1α'	Arg372	main-chain NH
	3γ	Lys377	side-chain NH	1α'	Asn317, Pro318	main-chain O
	3γ	Arg368	main-chain O	1γ	Lys31	main-chain NH
	3γ	Arg368	side-chain NH	1γ	Leu28	main-chain O
	1α'	Tyr369	side-chain OH	2α"	Glu35	side-chain O
	1α'	Lys375	side-chain NH	2α"	Leu28	main-chain O
	1α'	Arg368	main-chain O	2α"	Lys31	main-chain NH
	1α'	Lys377	side-chain NH	2β	Asn317	main-chain O
	1β'	Arg368	main-chain O	4β	Lys31	main-chain NH
	1β'	Met366[e]	main-chain O	4β	Lys31	side-chain NH

[*] No intra-subunit interactions are listed. All interactions are hydrogen bonds unless stated otherwise. Hydrogen-bonding interactions were identified using UCSF Chimera (*32*) with relaxed constraints of 0.4 Å and 20º, based on geometric criteria previously reported (*86*).

[a] Arabic numbers show the hexavalent pentamer from which the designated subunit originates, as labeled in Fig. 3.3C.

[b] Residues 347-358 form the KL β-hairpin, reference (*44*), Protein Data Bank entry 1SVA.

[c] Hydrophobic interface.

[d] Residues 367-378 form the KL β-hairpin for the β' subunit, 368-377 for the α' and γ subunits, reference (*50*) Protein Data Bank entry 1SIE.

[e] Residue Met366 in MPyV is on the N-terminal side of the K β-strand.

3.4 Discussion

3.4.1 Non-conserved C-terminus of VP1 may cause size variation in APV capsids

Our cryo-EM analysis of APV showed significant size variation among particles (38–65 nm) (Fig. 3.1). Size variation has been previously reported for APV isolated from infected birds or cell cultures (42–56 nm) (*91*-*95*) and for VLPs made from recombinant VP1 (52–58 nm) (*95*). Attempts to correct for relative size differences did not yield improved reconstructions, suggesting that size variation was accompanied by conformational variation.

The crucial role of the C-terminus in forming polyomavirus inter-capsomere connections suggests that the unusual size variation in APV is caused by the unique, truncated C-terminus of VP1 and not by any other intra- or inter-pentamer interaction, all of which are conserved (Figs. 3, S3). The C-terminal extension in APV VP1 is at least nine amino-acid residues shorter than the extension in any other polyomavirus—including those that infect other birds (Figs. 3A, S3). Compared to other polyomaviruses analyzed by cryo-EM by us (Fig. 3.1E) and others (e.g.(*9*, *45*, *47*, *81*)), the size variation of wild-type particles seems unique to APV. The biological role, if any, of APV size and conformational variation is unknown, but pliable capsids might allow more flexible and advantageous interactions with cellular receptors or may facilitate uncoating after cell entry. The variability may also mean that only certain particle classes are infectious.

3.4.2 Stabilizing function of the KL β-hairpin

Our results suggest that the KL β-hairpin acts to stabilize the polyomavirus capsid and is not a plug as previously suggested (*46*). The β-hairpins interact with two to four other VP1 subunits and connect two or three hexavalent pentamers (Table 3.2, Fig. 3.3D–F). Although the hairpin interactions are relatively few and weak, the C-termini of the α', β', and γ subunits do not simply

insert completely into the target pentamer (*96*) but also link additional pentamers. Mammalian polyomaviruses show density consistent with the interacting KL β-hairpins, but APV does not (Fig. 3.2, 3.3D,E). Observation of KL β-hairpin density at moderate resolutions suggests that these hairpins are stable entities and add rigidity to the overall virion structure. Indeed, genomic deletions of the KL β-hairpin in SV40 (residues 345–361) (*51*) and JCV (residues 339–354) (*52*) resulted in particles that exhibited significant size variation.

We liken the relatively weak KL β-hairpin interactions and the other, stronger C-terminal interactions to mechanical cotter and clevis pins, respectively (Fig. 3.3G). As a cotter pin stabilizes the clevis pin, the KL β-hairpin locks the stronger inter-capsomere interactions in place. Some *T*=3 viruses have components analogous to the β-hairpin of polyomavirus. For example, cowpea chlorotic mottle virus (CCMV) contains a β-hexamer motif made from six strands from six different subunits. Truncation of the CCMV β-hexamer did not prevent particle assembly, but the β motif does provide increased stability (*97*). Residues in a three-stranded, triangular-shaped β-annulus in Sesbania mosaic virus are not needed for particle assembly (*98*), but presumably stabilize the capsid. In polyomaviruses, CCMV, and Sesbania mosaic virus, respectively, the β-hairpin, β-hexamer, and β-annulus each seems to lock the capsid in a stable conformation.

3.4.3 Possible location of minor capsid protein VP4

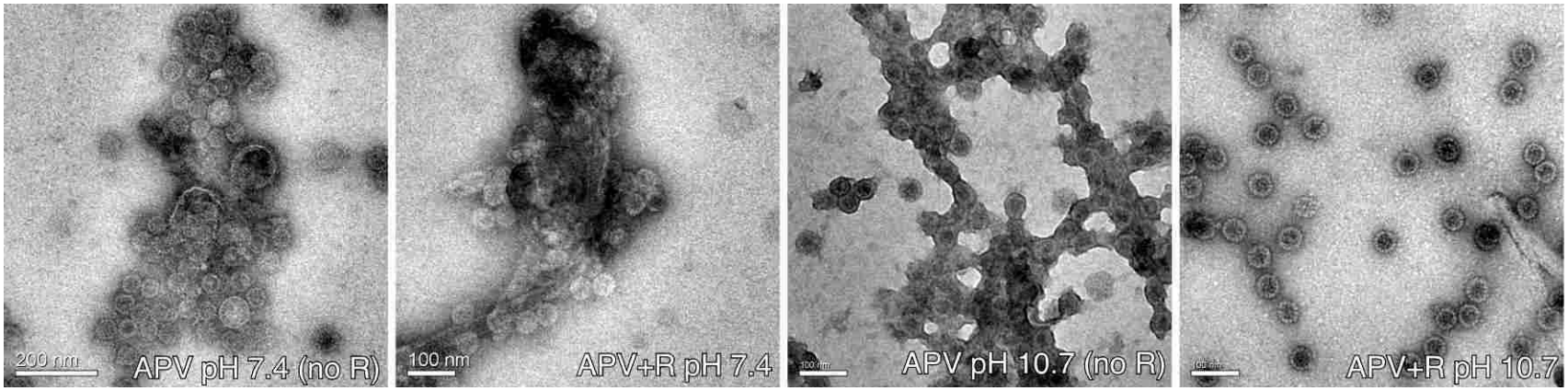
At the base of the axial cavities of the VP1 capsomeres, we observed larger plug-like densities in APV+R than in SV40 or JCV, and the plug-like density disappeared for JCV+R (Fig. 3.2). Minor capsid proteins were retained in APV and JCV capsids after arginine treatment (Fig. 3.4). At least three possibilities might explain these phenomena: 1) VP4 is bound in the axial channel adjacent to VP2/3 and shields these minor proteins from the dissociating effects of arginine; 2)

VP1, VP2/3, or both are more ordered and stable in APV+R than in mammalian polyomaviruses and VP4 is disordered in the icosahedrally averaged structure; and 3) as in JCV, APV VP2/3 dissociates from the chamber in the presence of arginine, leaving VP4 to form the complete plug. The sequences of VP2 and VP3 are comparable between bird and mammalian polyomaviruses, with conserved and non-conserved residues in the same regions (Fig. S3.5). Chen et al. (*54*) observed that the C-terminal portion of VP2/3 forms an α-helix that interacts specifically with VP1 and postulated that this interaction is conserved among all polyomaviruses. Therefore, if the only ordered portion of VP2/3 is the conserved VP1-binding region, then the extra density in APV+R would be something other than VP2/3, or the entire plug-like density would be VP4.

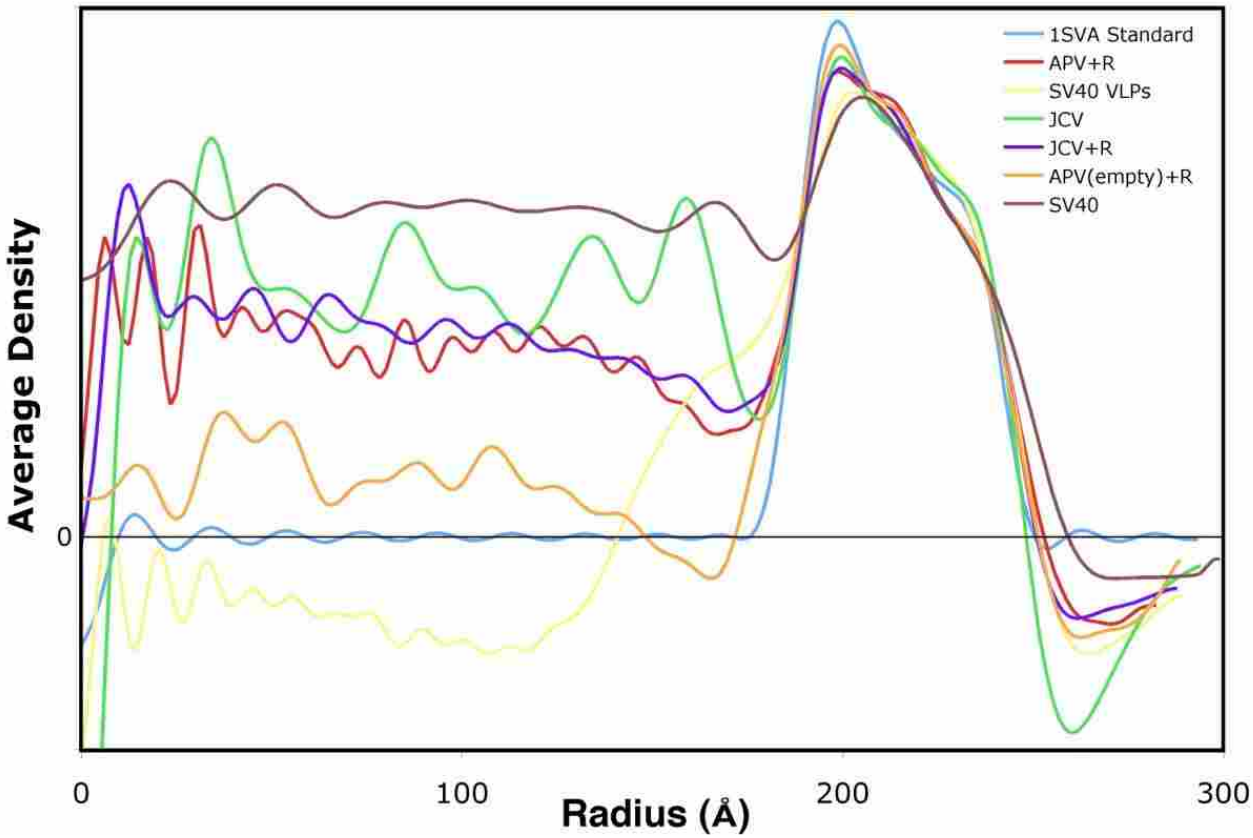
Placement of VP4 between VP1 and the genome is consistent with the demonstrated binding of VP4 to both VP1 and dsDNA (*61*). A VP1-VP4 interaction was also shown by their co-localization in the nucleus (*64*). VP4 is required for proper APV assembly, as deletion of VP4 resulted in a higher percentage of mis-assembled particles (*23*). Recombinant baculovirus expression of only VP1–3 was insufficient to form APV VLPs (*99*), whereas expression of mammalian polyomavirus VP1–3 does form VLPs (*100*). Although no proof of the VP4 location is possible from our data, a position at the base of the axial cavity, in proximity to VP2/3, would be consistent with the proposed function of VP4 in APV genome packaging and virion assembly (*23, 61*).

3.5 Appendix

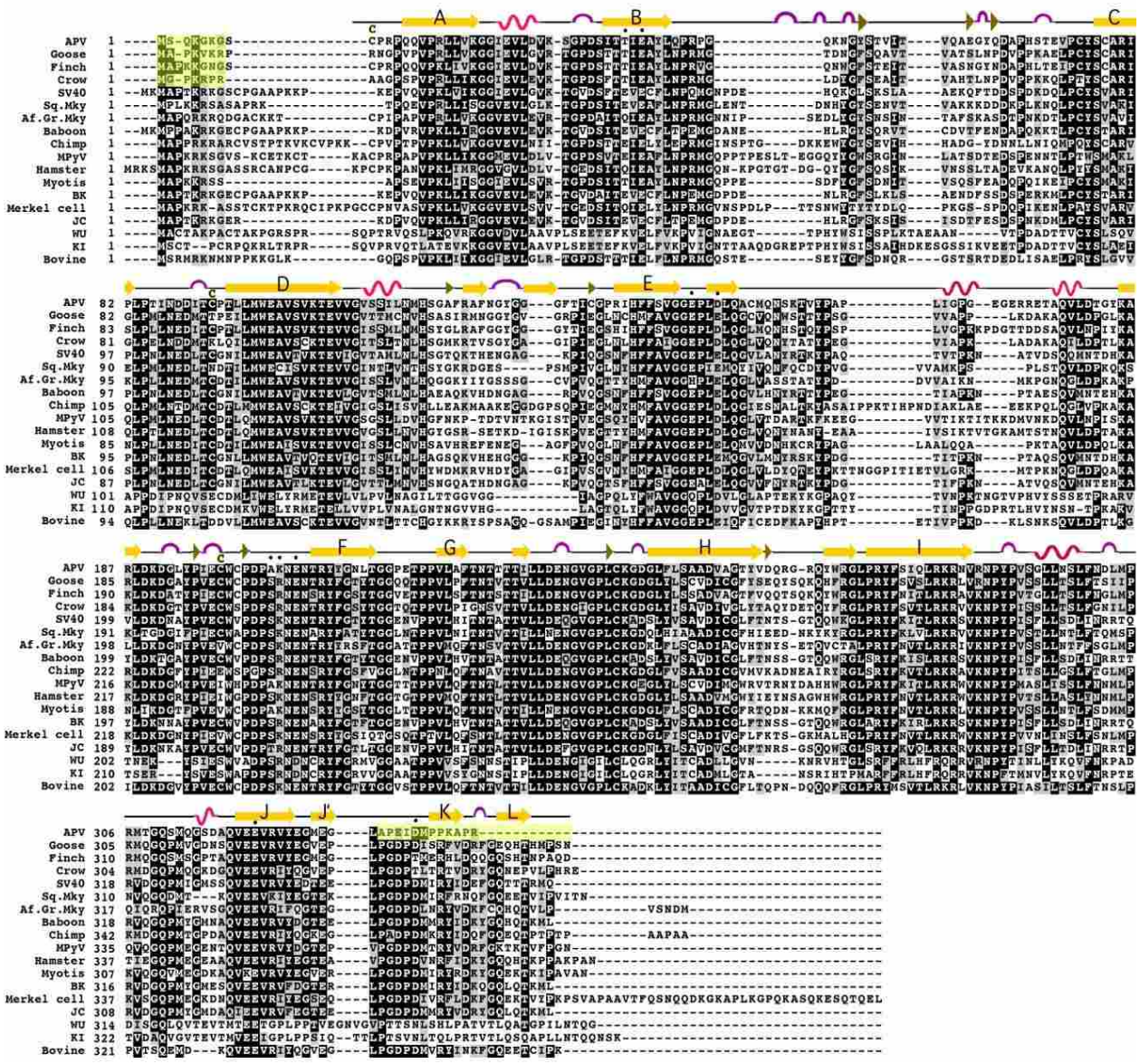
3.5.1 Supplementary figures



Supplementary Figure 3.1. (Fig. S3.1) Negatively stained APV particles under different solution conditions. APV particles aggregate under mild buffer conditions (APV pH 7.4 (no R)). APV particles in a de-aggregated state when treated with 250 mM L-arginine (APV+R pH 10.7). Particles re-aggregate if pH is reverted to 7.4 (APV+R pH 7.4). Adjusting pH to 10.7 (APV pH 10.7 (no R)) is insufficient to de-aggregate particles. Therefore, the combination of arginine and high pH is necessary to make sample suitable for cryo-EM. We also treated with arginine and then dialyzed out the arginine. The particles aggregated and were multi-sized as before (data not shown). This latter experiment suggests that arginine treatment does not produce irreversible changes in the capsid.



Supplementary Figure 3.2. (Fig. S3.2) Calibrated spherically averaged density profiles of computed reconstructions. Profiles were calibrated against SV40 standard (Stehle et al., 1996) (PDB ID, 1SVA). Background density is zero. The capsid extends from radius 174 Å to 255 Å. Note that the empty structures (1SVA, SV40 VLP, and APV(empty)+R) have average inner densities at or near zero.

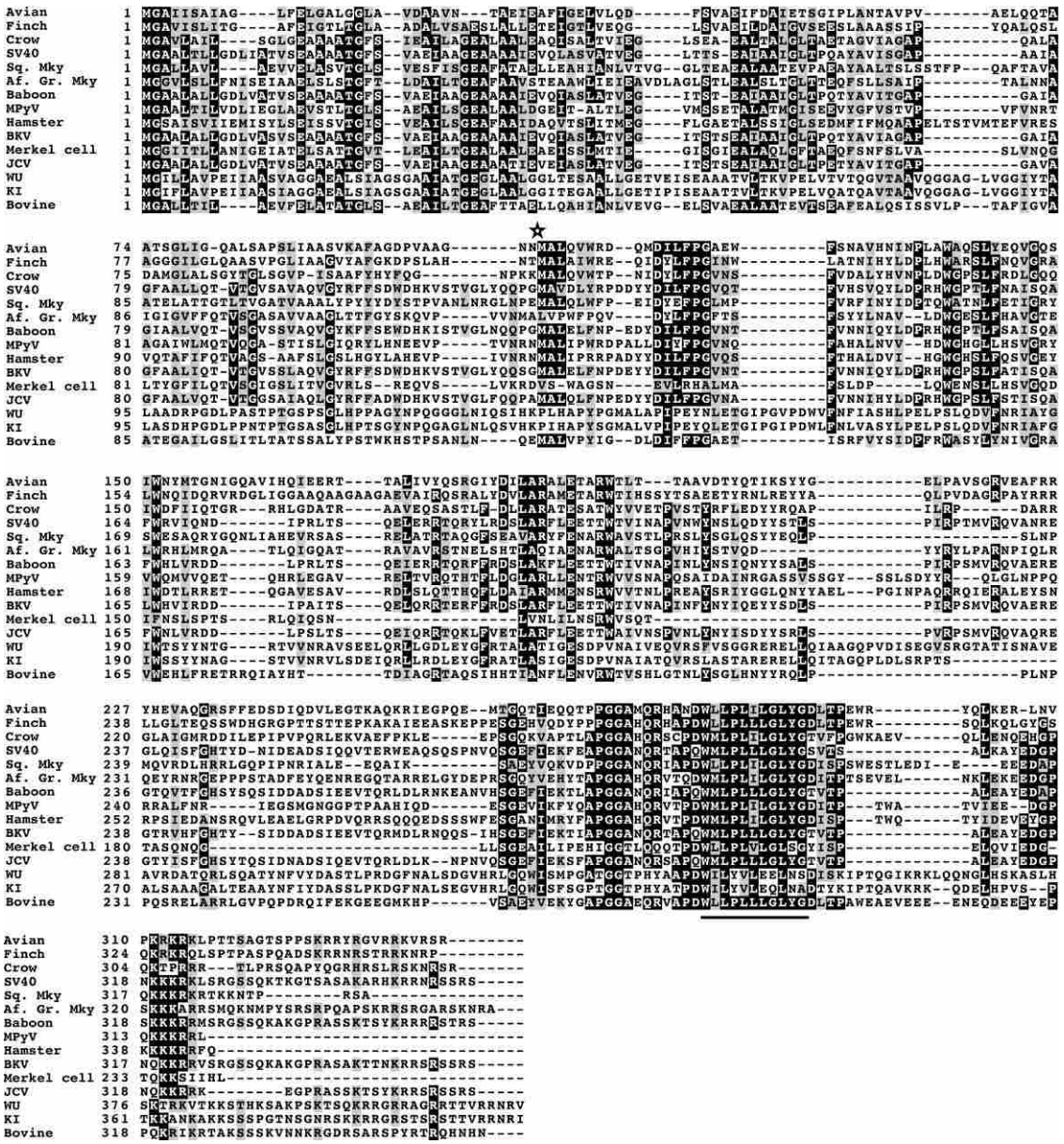


Supplementary Figure 3.3. (Fig. S3.3) VP1 sequence alignment among 18 polyomavirus variants. Top four rows are bird sequences. Secondary structure of MPyV depicted above sequence alignment with β-strands A-L labeled. Highlighted regions at N- and C- termini emphasize differences in nuclear localization signal and KL β-hairpin, respectively. White letters in black boxes: >50% conservation; black letters in gray boxes: residues with conserved properties (e.g. basic); black letters in white boxes: non-conserved residues. Dots, putative Ca^{2+} binding sites (Stehle et al., 1996). Bold 'C' (below secondary structure), disulfide-forming residues (Stehle et al., 1996) and (Jao et al., 1999). Excluding KI and WI polyomaviruses, pairwise alignment scores (percent identity) were 56-73% among bird sequences, 41-62% for birds vs. mammalian viruses, 47-83% among mammalian viruses, KI and WU were 67% identical to each other but were only 23-29% identical to all other bird and mammalian viruses. In addition to the non-conserved residues of APV in the KL-hairpin region, KI and WU polyomaviruses contain non-conserved residues in this region, but their sequences are not truncated as in APV. Note that Merkel cell polyomavirus is at least 34 residues longer than any other polyomavirus. Baboon = baboon polyomavirus 2; Sq. Mky = squirrel monkey; Af. Gr. Mky = African green monkey.

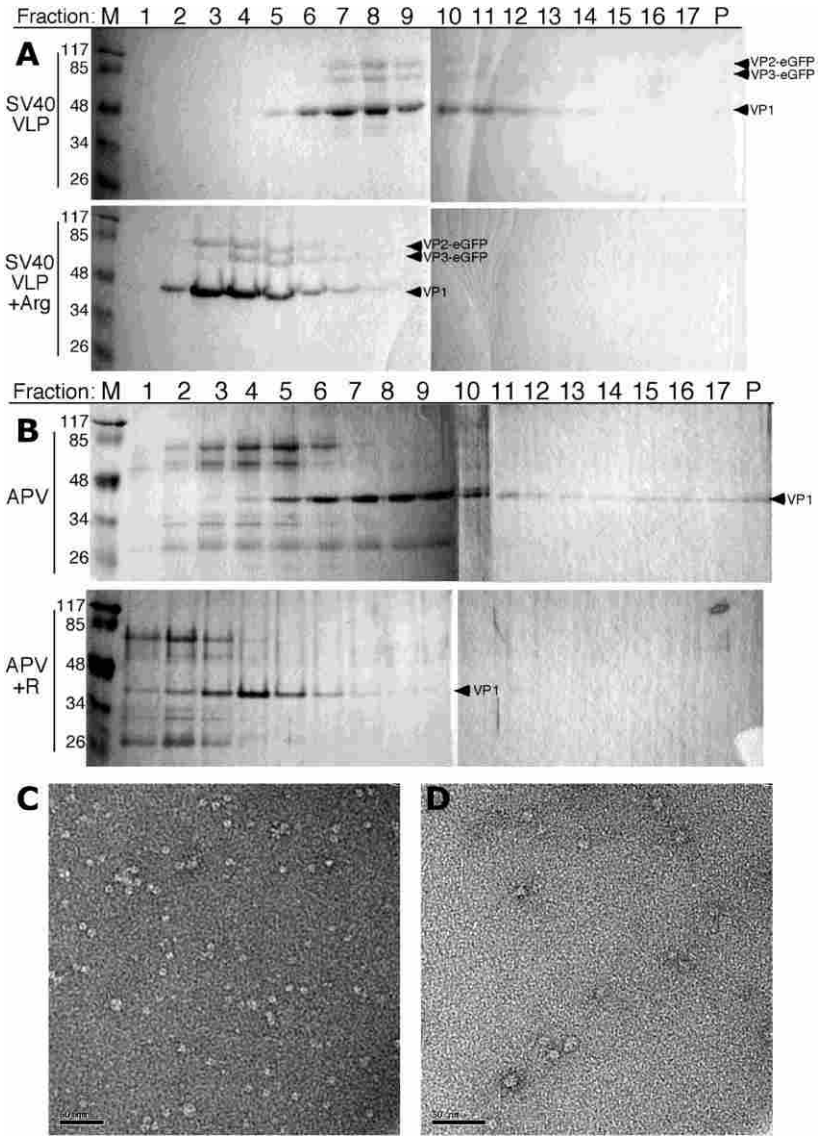
Supplementary Figure 3.4. (Fig. S3.4) Scores for pairwise alignments between polyomavirus VP1 amino-acid sequences.Yellow – mammal-mammal alignments; Blue – bird-bird alignments; Green – bird-mammal alignments, bolded cells emphasize highest and lowest alignment scores among the bird-mammal scores, excluding alignments with human WU and KI virus.

APV, avian polyomavirus; GHPyV, goose hemorrhagic polyomavirus; FPyV, finch polyomavirus; CPyV, crow polyomavirus; SV40, simian virus 40; SMPyV, squirrel monkey polyomavirus; LPyV, African green monkey polyomavirus; BaPyV, baboon polyomavirus; ChPyV, chimpanzee polyomavirus; MPyV, murine polyomavirus; HPyV, hamster polyomavirus; Myotis, myotis polyomavirus; BKV, BK virus; MCPyV, Merkel cell polyomavirus; JCV, JC virus; WU, WU virus; KI, KI virus.

	GHPyV	FPyV	CPyV	SV40	SMPyV	LPyV	BaPyV	ChPyV	MPyV	HPyV	Myotis	BKV	MCPyV	JCV	WU	KI	BPyV
APV	60	73	56	52	49	51	53	52	**41**	54	58	51	52	52	29	27	49
GHPyV	x	63	70	56	58	57	57	55	57	57	**62**	53	55	56	27	24	58
FPyV	x	x	59	55	52	53	55	55	55	60	60	55	55	54	29	27	52
CPyV	x	x	x	56	55	59	56	55	56	56	58	54	56	55	26	26	57
SV40	x	x	x	x	52	51	78	48	54	52	53	81	49	76	27	26	50
SMPyV	x	x	x	x	x	53	50	52	50	52	59	49	53	50	26	24	54
LPyV	x	x	x	x	x	x	51	54	53	58	58	51	57	52	27	25	51
BaPyV	x	x	x	x	x	x	x	50	53	50	54	83	49	80	26	25	47
ChPyV	x	x	x	x	x	x	x	x	52	52	52	51	54	48	25	23	52
MPyV	x	x	x	x	x	x	x	x	x	65	56	53	53	53	27	26	53
HPyV	x	x	x	x	x	x	x	x	x	x	58	50	59	54	28	24	52
Myotis	x	x	x	x	x	x	x	x	x	x	x	52	60	54	24	24	57
BKV	x	x	x	x	x	x	x	x	x	x	x	x	48	78	25	26	47
MCPyV	x	x	x	x	x	x	x	x	x	x	x	x	x	48	27	23	52
JCV	x	x	x	x	x	x	x	x	x	x	x	x	x	x	29	28	49
WU	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	67	24
KI	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	25



Supplementary Figure 3.5. (Fig. S3.5) VP2 and VP3 sequence alignment among 15 polyomavirus variants. The first three rows are bird sequences. White letters in black boxes: >50% conservation; black letters in gray boxes: residues with conserved properties (basic, acidic, hydrophobic, etc.); black letters in white boxes: non-conserved residues. Black bar denotes conserved VP1-binding helix. Star denotes the beginning of VP3 sequence. Baboon = baboon polyomavirus 2; Sq. Mky = squirrel monkey; Af. Gr. Mky = African green monkey. Note that conserved regions are found in both mammalian and bird polyomaviruses.



Supplementary Figure 3.6. (Fig. S3.6) L-Arginine, DTT, and EDTA disrupt VP1 pentamers of SV40 (A) and APV (B–D). SV40 VLPs and APV were dissociated into pentameric subunits by use of 10 mM EDTA, 10 mM DTT in buffer (100 mM NaCl, 5% glycerol, 20 mM Tris-HCl pH 7.5) with and without L-arginine. Dissociated SV40 samples were loaded onto 5%-20% sucrose gradients with and without arginine and ultracentrifuged in a SW55Ti rotor for 3 hours at 55k rpm, 4 ºC. A total of 18 fractions were collected after ultracentrifugation; each fraction was loaded onto a 15% SDS-PAGE gel. Without arginine, VP2 and VP3 co-migrated with VP1 pentamers approximately 1/2 down the length of the tube (A, top row). However, the addition of arginine caused proteins to stay in the top portions of the gradient (A, bottom row). Similar results were seen with APV, which was loaded on a 5–40% sucrose gradient with (B, bottom row) and without (B, top row) arginine and ultracentrifuged in a SW55Ti rotor for 3 hours at 50k rpm, 4 ºC. Fractions were taken from the top (1) to bottom (17,P) of the centrifuge tube. M, marker (standard shown in kDa) lane; P, pellet. SV40 minor capsid proteins VP2 and VP3 were fusion proteins attached to enhanced green fluorescent protein (eGFP), and therefore, had a higher mass than VP1. Note: In the APV gels (B), the bands between 48 and 85 kDa are regularly observed in APV SDS-PAGE gels and most likely represent remaining aggregates of VP1 or other impurities. The bands less than 34 kDa are likely minor capsid proteins VP4 and VP3, as well as histones. Compare to Fig. 3.4A,B. (C,D) Negatively stained specimens (C) APV dissociated with 10 mM DTT and 10 mM EDTA produces consistently sized pentamers, e.g. arrow. D) the presence of L-arginine in the same dissociation buffer causes non-uniformly sized aggregates and no discernable VP1 pentamers. Bars, 50 nm.

3.5.2 Genbank Accession Numbers

VP1: APV, BAH22404; MPyV, AAA93242; SV40, AF180737; BKV, YP_717939; JCV, BAC22906; finch polyomavirus, ABB04271; goose hemorrhagic polyomavirus, AQ91620; baboon polyomavirus (type 2), YP_406554; African green monkey polyomavirus, NP_848007; squirrel monkey polyomavirus, YP_001531348; crow polyomavirus, YP_529827; hamster polyomavirus, CAA06802; bovine polyomavirus, NP_040787; WU polyomavirus, ACR43499; KI polyomavirus, ACR43495; myotis polyomavirus, ACI16488; Merkel cell polyomavirus, ACN44192

VP2: APV, BAH22430.1; finch polyomavirus, YP_529831.1; crow polyomavirus, YP_529825.1; SV40, AAK01715.1; squirrel monkey polyomavirus, YP_001531346.1; African green monkey polyomavirus, NP_848005.1; baboon polyomavirus (type 2), YP_406552.1; MPyV, P03097.3; hamster polyomavirus, NP_056734.1; BKV, BAE53664.1; Merkel cell polyomavirus, ACL31698.1; JCV, BAC15711.1; WU polyomavirus, YP_001285485.1; KI polyomavirus, A3R4N1.1; bovine polyomavirus, NP_040785.1

APV Genomes: BFDV, AF118150; BFDV-1, AF241168; BFDV-4, AF241169; BFDV-5, AF241170; APV1, AB453159; APV2, AB453160; APV3, AB453161; APV4, AB453162; APV5, AB453163; APV6, AB453164; APV7, AB453165.

3.5.3 Acknowledgments

Co-authors in this study include Dirk Enderlein, Christian D.S. Nelson, Weston S. Carter, Masaaki Kawano, Li Xing, Robert D. Swenson, Norman H. Olson, Timothy S. Baker, R. Holland Cheng, Walter J. Atwood, Reimar Johne, and David M. Belnap.

This research was supported by a Brigham Young University Mentoring Environment Grant and other BYU institutional funds to DMB; BYU Cancer Research Center fellowships, a BYU Graduate Research Fellowship, a Roland K. Robins Graduate Research Fellowship, and a Jennie R. Swensen Graduate Research Fellowship to PSS; a grant from the Deutsche Forschungsgesellschaft (project JO 369/3–2) to DE; Ruth L. Kirschstein National Research Service Award (F32) 1F32NS070687 to CDSN; National Institutes of Health grant R37-GM033050 to TSB; University of California Discovery Grant 05-10505, NIH grant AI066287-01A1, and National Cancer Institute Pilot Grant (CA093373-06S1) to RHC; and NIH grant R01NS043097 to WJA. The BYU Fulton Supercomputing Lab provided computer resources.

We thank Matthew Masner, Jenny Shen, Daniel Shen, and David Eng for help with image processing; Kristin Pande for laboratory assistance; Rebecca Page for help with size-exclusion experiments; Irene Larsen for early work on this project; Jeffrey Farrer and Michael Standing for microscopy support; J. Bernard Heymann for software support; J. Zachary Porterfield for suggesting arginine treatment to deaggregate APV; Hiroshi Handa for SV40 VLP expression vectors; Adam Zlotnick for helpful discussions; our peer reviewers for helpful suggestions on the manuscript; Eduardo Sanz-Garcia for insightful discussions and help with computer programs; and Hermann Müller for ongoing encouragement of this project. The UCSF Chimera package (http://www.cgl.ucsf.edu/chimera) is supported by NIH P41 RR-01081. The APV homology model is available as entry 3IYS in the Protein Data Bank.

Chapter 4

Conformational variance of satellite tobacco mosaic virus capsid around a stable RNA scaffold

4.1 Introduction

Satellite tobacco mosaic virus (STMV) is a small, $T = 1$ icosahedral particle approximately 17 nm in diameter (*7*). The capsid is composed of 60 identical copies of the capsid protein and encases a 1,059 nucleotide, (+) single-stranded, RNA genome. The total molecular mass of STMV is approximately 1.3×10^6 Da. Infection of tobacco plants and other members of the family Solanaceae by STMV is only possible through co-infection with a tobamovirus (*101*), which may heighten symptoms compared to infection by tobamovirus alone (*102*).

The molecular structure of STMV has been resolved by X-ray crystallography (*7*). The capsid protein arranges in a β-barrel jellyroll fold with extended amino-terminal arms that interact with the RNA. Capsid protein dimers are presumed to be the assembly unit of the capsid due to the large amount of hydrophobic and hydrogen-bond interactions at the dyad axes. Conversely, capsid protein interactions are fewer at the three- and five-fold axes.

The RNA core of STMV is inherently stable in the absence of capsid proteins (*103, 104*). The crystal structure of STMV revealed partially-ordered RNA throughout the capsid interior (*7*). The RNA arranges inside the capsid as 9-base double-helical segments at every dyad axis. These 30 helices account for about 50% of the whole genome. The capsid-RNA interface is lined with amino acids that interact with the nucleic acid backbone via hydrogen bonding and salt bridges. Extracted RNA cores maintain a globular shape with a diameter corresponding to the

interior of the capsid and are able to unfold and fold reversibly (*103*). Molecular dynamic simulations suggest that the viral genome plays an important role in capsid assembly (*105*) and stability (*106*). These observations led to the assembly model where RNA must be completely folded before the binding of capsid proteins (*106*). Therefore, STMV assembly seems to be driven by the RNA core rather than by the tendency of capsid proteins to adopt an icosahedral shape. In fact, capsids without RNA have rarely been observed (*106*).

Here, we present solution-state structures of STMV at pH 7 and 10, solved by cryo-EM, which revealed different capsid arrangements than the reported X-ray crystal structure of the virion (*7*). Rigid-body fitting of the crystal coordinates into the cryo-EM density maps accounted for the conformational differences between the crystal and cryo-EM structures. Yet, the reconstructions contained conserved densities that matched well to the icosahedrally symmetric RNA core found in the crystal structure. The observed capsid variation supports the notion of flexibility in capsid protein arrangement around a stable RNA scaffold.

4.2 Materials and methods

4.2.1 Cryo-EM and image processing

Purified STMV was either kept in pure water or diluted to a final concentration of 1.3 mg/mL in Buffer A (10 mM Tris-Cl, 50 mM NaCl, 0.01 mM $CaCl_2$, pH 7.4). 3.5 μL of sample was applied to a holey carbon copper grid with a thin layer of carbon applied over C-flat holey grids (Protochips, Raleigh, North Carolina, USA). The grids were vitrified in liquid ethane using a Vitrobot (FEI) vitrification device (5 second blot) at 4.4°C and 100% relative humidity. Specimens were transferred to a Gatan 626 cryoholder (Pleasanton, California, USA) cooled with liquid nitrogen. Cryo-EM images were recorded at 200 keV and at

59,000x magnification via low-dose methods at objective lens underfocus levels ranging from 0.5–4.0 μm on Kodak SO-163 film in an FEI Tecnai F30 transmission electron microscope (Hillsboro, Oregon, USA).

Electron micrographs were digitized on a Nikon Super Coolscan 9000 ED scanner in 16-bit, positive mode. Particle images were extracted from electron micrographs using *X3DPREPROCESS* (*74*). We determined and corrected for contrast transfer function and decay parameters using Bsoft (*bshow* and *bctf* functions) (*75*). Origins and orientations of the extracted particles were determined via the model-based technique of *PFT2* (*76*), which was adapted to use phase and amplitude information in orientation selection (*31*). A 15-Å electron density map of STMV was generated from the crystal structure (PDB ID: 1A34 (*7*)) to initiate model-based reconstruction. Final reconstructions were calculated using *EM3DR2* (http://people.chem.byu.edu/belnap/pft3dr). Bsoft (*blocres* function) (*75*) was used to estimate quasi-local-resolution by comparing 60 voxel (2.13×10^5 Å^3) sub-volumes centered at each voxel in the capsid portion of the final STMV reconstruction.

For size calibration purposes, we used poliovirus (PV, 160S state (*107*)) as an internal standard. PV particles were mixed with STMV in Buffer A. The combined sample was prepared for cryo-EM, recorded, and digitized as described above. 294 PV particles were extracted from the micrographs, which produced a 20-Å reconstruction. A previously calibrated map of 160S PV was used to determine the size of the PV reconstruction, and, hence, the sampling size of the STMV reconstruction. The sampling sizes of all other STMV reconstructions, from images taken at the same buffering conditions and voltage levels, were adjusted by comparing the reconstructions to the calibrated STMV map.

4.2.2 Structure modeling and analysis

Rigid-body fitting of STMV coordinates (PDB ID: 1A34 (*7*)) into electron density maps was accomplished using the *Situs* package (*108*). This tool performs an exhaustive search in 6D space (3 translational and 3 rotational degrees of freedom) for the most favorable fitting of the crystal coordinates into the electron density map. Crystal contacts and atomic clashing were determined with *UCSF Chimera* (*32*). Flexible loops were modeled using *MODELLER* (*109*).

4.3 Results and discussion

4.3.1 Conformational changes of STMV capsids

We computed a reconstruction of STMV using 79,395 particle images (94,850 total, 83.7% selected) extracted from 54 focal pairs of micrographs. Despite using a 15-Å map of the STMV crystal structure (PDB ID: 1A34 (*7*)) as a starting reference model, the iterative refinement process of computing the 3D reconstruction resulted in a consistently divergent 14-Å structure (Fig. 4.1).

Although the radius of the STMV reconstruction was identical to the radius of the STMV crystal structure, their capsid structures are clearly dissimilar (Fig. 4.1, 4.2). Differences between the cryo-EM and crystal structures were prominent at the five-fold, three-fold, and two-fold axes of the icosahedral shell. At the five-fold axes, the cavity formed by the convergence of five protein subunits is present in both structures, but the star shape formed at each vertex is rotated counterclockwise in the cryo-EM reconstruction. This rotation results in a redistribution of density throughout the capsid without affecting the overall radius. The three-fold axis from the crystal structure consists of shallow grooves surrounded by bumps formed around the five-fold axis. In the cryo-EM reconstruction, the three-fold axes are filled with densities that converge at

each axis such that the grooves are lost. Instead, the three-fold axis of the cryo-EM structure is considerably smoother than the crystal structure map at the same location and resolution. The opposite was observed at the two-fold axis where a ridge of density in the crystal structure map is in contrast to the deep groove seen in the reconstruction.

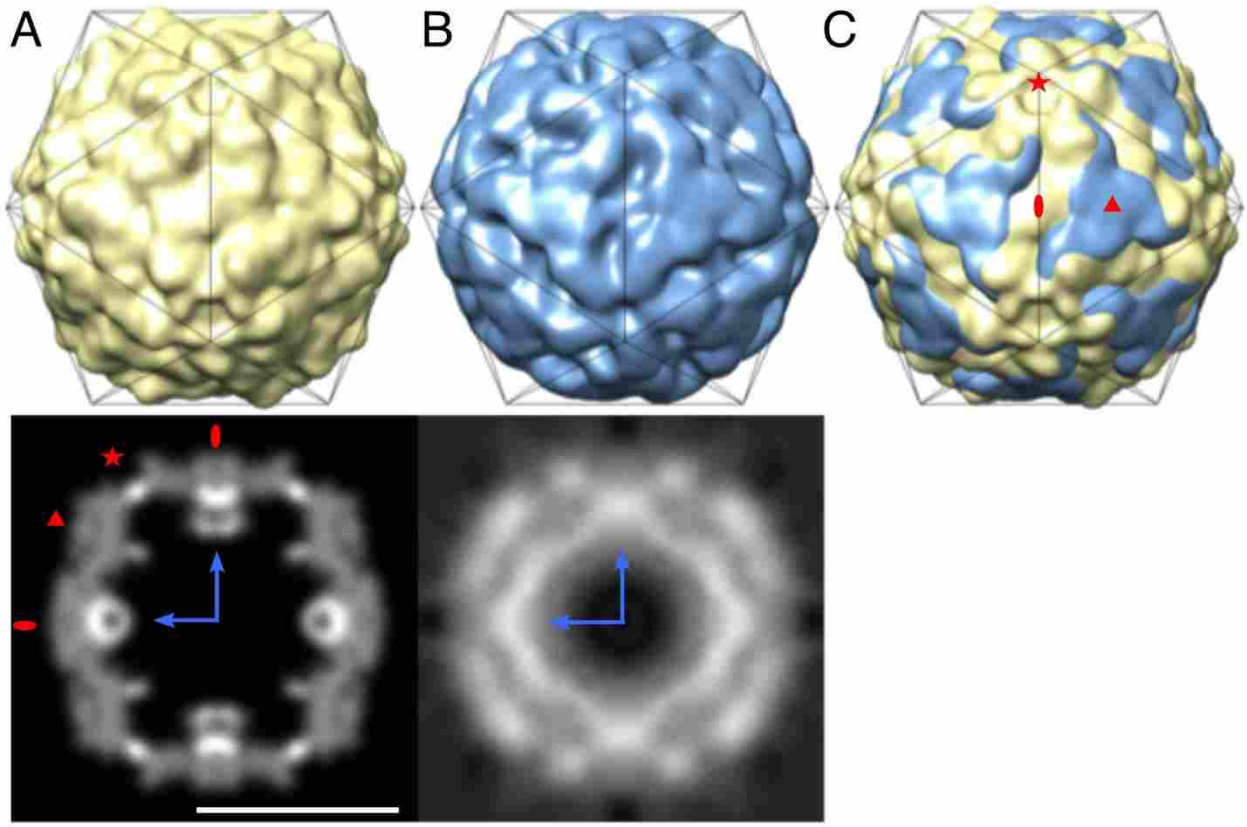


Figure 4.1. Differences between STMV crystal and solution structures. (A) Electron density map computed from crystal coordinates of STMV. (B) Density map of STMV computed from cryo-EM and single-particle reconstruction methods. (C) Superimposed structures from (A) and (B) show the differences between the crystal and cryo-EM structures. Despite the differences in capsid structures, portions of the RNA core appear conserved (arrows). The icosahedral two-fold (oval), three-fold (triangle), and five-fold (star) axes are labeled. Top row, surface rendering at 1.0-σ contour level. Bottom row, central slices. Scale bar, 10 nm.

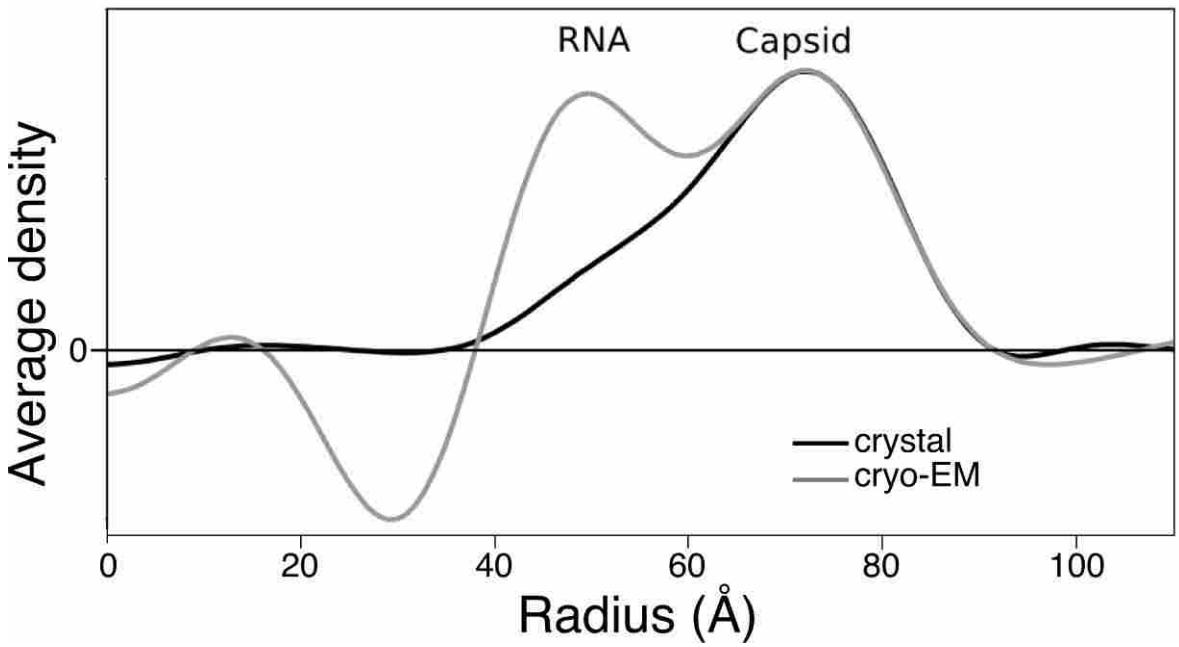


Figure 4.2. Spherically-averaged density profiles for density maps of STMV derived from crystallography (PDB 1A34, (*7*)) and cryo-EM methods. Cryo-EM structures were calibrated using 160S poliovirus as an internal standard. The overlapping of capsid peaks is an indication of conserved sizes between crystallography and cryo-EM structures. Differences in the RNA peaks correspond to the extra densities of RNA that were not visible in the crystal structure (see Fig. 4.1).

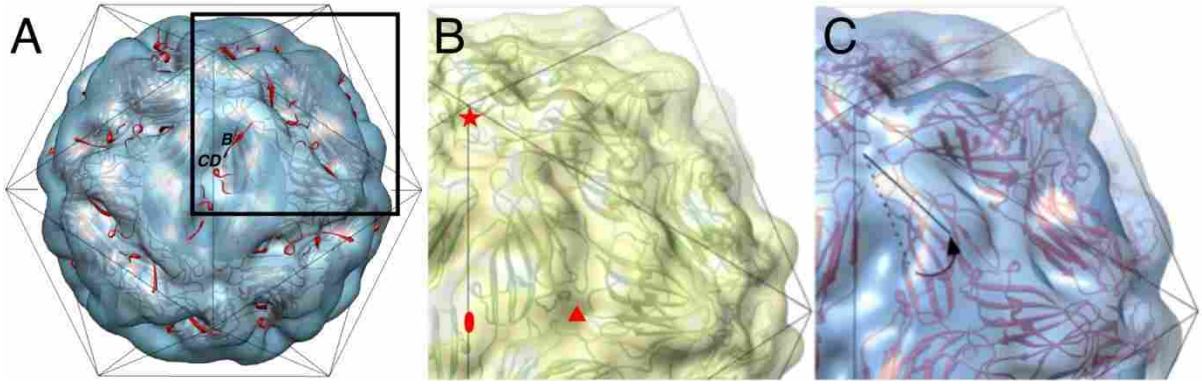


Figure 4.3. STMV crystal coordinates (ribbons) fitted into the cryo-EM reconstruction (blue). (A) Original crystal coordinates overlaid against the cryo-EM reconstruction. A B-strand and CD loop is labeled. Box outlines zoomed portions for (B) and (C). (B) Original crystal coordinates overlaid against its density map (yellow). Icosahedral axes labeled as in Fig. 4.1. (C) STMV crystal coordinates fitted into the cryo-EM reconstruction. The starting position of the crystal coordinates (dashed line) was shifted and rotated about the five-fold axis (arrow) until the coordinates matched the cryo-EM density map (solid line).

To characterize the structural differences between the crystal and cryo-EM structures, we fitted the atomic coordinates of the STMV β-jellyroll fold into the cryo-EM reconstruction using rigid-body fitting techniques. Prior to fitting, the most prominent segment to protrude from the density map was the loop formed between strands C and D (CD) and a portion of strand B (Fig. 4.3A). We achieved a favorable fitting by shifting and rotating the capsid protein about the five-fold axis (Fig. 4.3C). Fitting the core jellyroll fold resulted in only 3.5% (70 of 1970) of the Cα atoms protruding out of the density. The fitting did not produce any steric clashing among protein subunits. The fitting accounted for the rotated star-shaped formation about the five-fold axis, increased densities at the three-fold axis, and showed absence of densities at the two-fold axis.

Rigid-body fitting of the capsid protein resulted in the formation of deep grooves at the two-fold axes, as seen in the cryo-EM structure (Fig. 4.4). These conspicuous features indicate that capsid protein interactions are minimal at the dyad axes. However, the axial point is marked by a noticeable bulge that stems out of the groove (Fig. 4.4B). Our fitting did not reveal any noticeable interactions at the two-fold axes that could explain the bulging densities. However, the large, flexible loop formed between beta-strands E and F (EF) could presumably be modeled into the bulge. We used *MODELLER* to determine the loop conformations with the lowest discrete optimized protein energy value. This conformation, and others with similar energy values, coincided with the two-fold axis bulge (Fig. 4.4C). Potential interactions between two EF loops were revealed after icosahedral symmetry was applied to the remodeled loop. The loop consists of residues that could form salt bridges (Lys101, Asp104, Glu107, Glu108) and other hydrogen bonds (Thr102, Ser105, Thr106) with another EF loop. Nevertheless, the total number of inter-subunit interactions at the two-fold axis is severely diminished after fitting. In the crystal

structure, at least 18 unique hydrogen bonds stabilize the dimer, not including additional water-mediated hydrogen bonds (*7*).

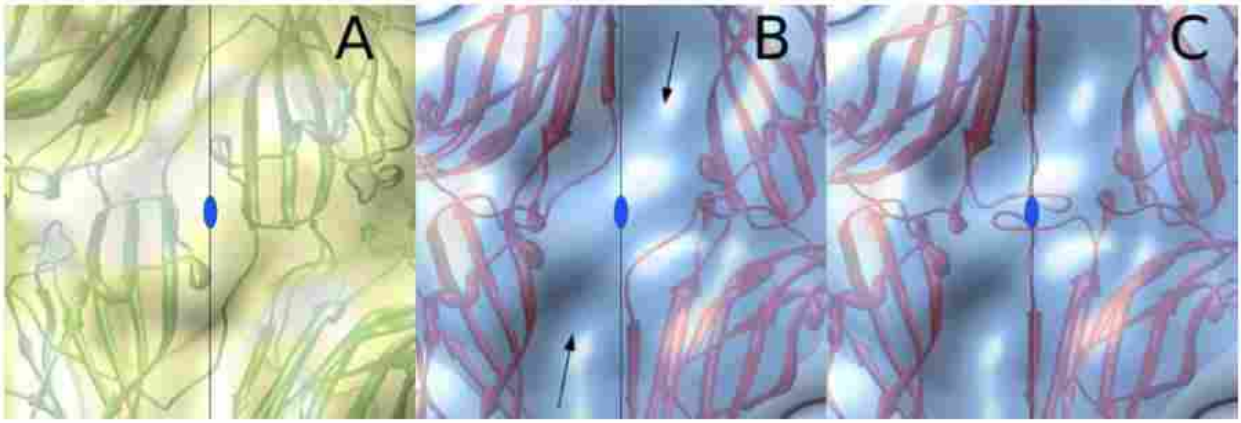


Figure 4.4. Differences between the STMV crystal and cryo-EM structures about the two-fold axis (oval). (A) Original STMV coordinates (PDB 1A34, (*7*)) and accompanying density map. (B) Rigid-body fitting of the crystal structure into the cryo-EM reconstruction. The fitting explains the lack of density around the two-fold axis (indicated by arrows). (C) The EF loop was remodeled and accounts for the bulge seen at the two-fold axis. Views are from outside the particle.

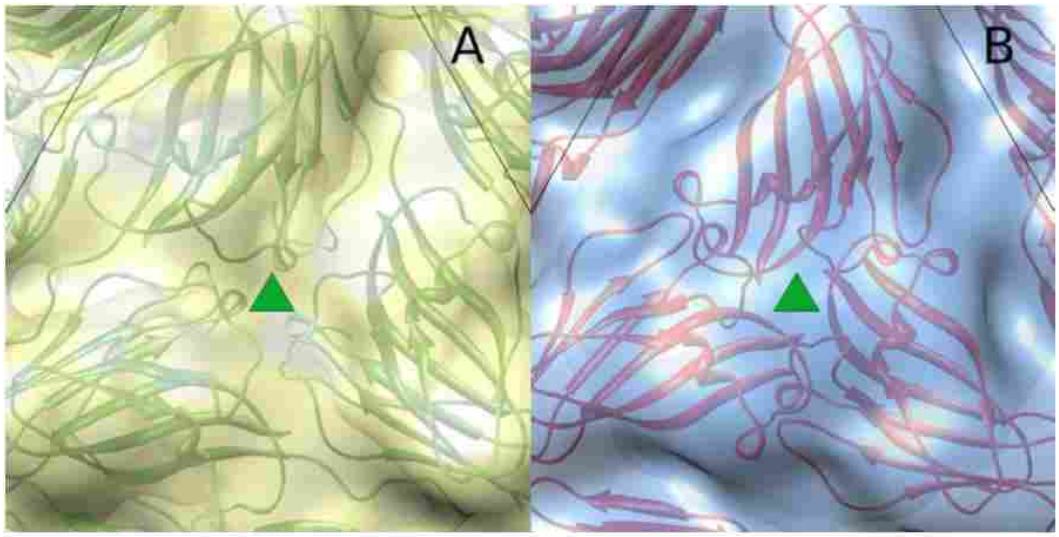


Figure 4.5. Differences between the STMV crystal and cryo-EM structures about the three-fold axis (triangle). (A) Original STMV coordinates (PDB 1A34, (*7*)) and accompanying density map. (B) Rigid-body fitting of the crystal structure into the cryo-EM density map accounts for increased densities at the three-fold axis.

The fitting resulted in the merging of three capsid protein subunits at the three-fold axes (Fig. 4.5). The convergence of the proteins suggests that more intermolecular interactions occur at the three-fold axes than are seen in the crystal structure. Indeed, a closer investigation of plausible interactions at the three-fold axes suggests increased hydrogen-bond interactions and hydrophobic interfaces (Fig. 4.6). The interactions at the three-fold axis occur largely among flexible loops, which may be further remodeled to produce more favorable interactions. In all, the collective differences between the two structures at the five-, three-, and two-fold axes suggest a conformational switch from a two-fold-dominated arrangement in the crystal structure to a three-fold-dominated arrangement in the cryo-EM reconstruction.

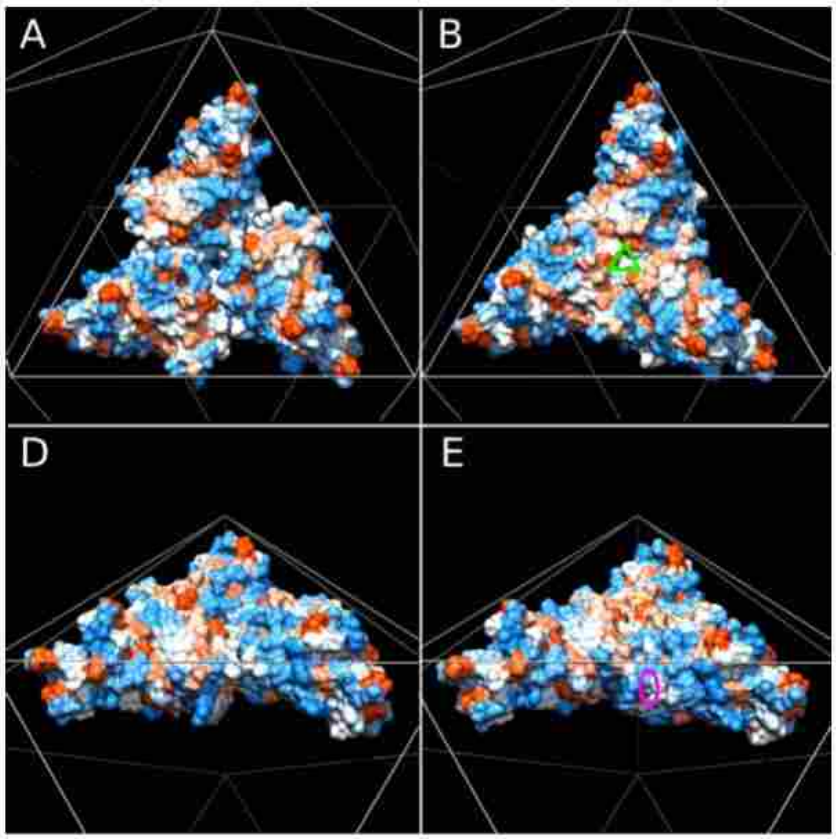


Figure 4.6. STMV hydrophobic interactions at the three-fold axis. (A) Molecular surface representation of three STMV capsid proteins in the crystal structure configuration (PDB 1A34, (*7*)). Residues are colored according to hydrophobicity (red, non-polar; blue, polar; intermediate values in white). (B) Fitting the coordinates into the cryo-EM map resulted in the formation of a hydrophobic patch at the three-fold axis (triangle). (C and D) Side views of (A) and (B), respectively. Polar residues are found at the two-fold axis (oval).

Although the fitting relied on rotational and translational adjustments about the five-fold axis, the subunit interactions about the axis appear unchanged. The shallow cavity formed at the five-fold axis is present in both structures, indicating favorable interactions formed among the five subunits at that location. In the crystal structure, the convergence of the five subunits at the icosahedral vertices is marked by a network of asparagine side chains (Asn115 and Asn117), which form a ten-ligand cage about the axis. The cage was shown to surround a single sulfate ion, which presumably facilitates hydrogen-bonding with the asparagine cage (*7*). The buffers used in this study lacked sulfates and phosphates, suggesting that large anions contribute to the crystal structure conformation. Nevertheless, the caged anion is not requisite for the overall structural integrity, as previously suggested (*7*) and supported by our fitting. In the new conformation, hydrogen bonding interactions among the ten asparagines (~3Å apart) may be sufficient to stabilize the interactions among the five subunits.

The modest resolution of our reconstruction (14-Å) was insufficient to accurately fit the coordinates of the amino-terminal "arm" into the density map. Nevertheless, the rigid-body fitting of the major fold maintained the general direction of the arm into the capsid interior. The crystal structure of STMV revealed that several basic residues of the arm interact with the RNA core in a nonsequence-specific manner. Our fitting of the coordinates upholds the tethering nature of the arm, assuming that it is somewhat flexible.

We investigated whether the conformation of the STMV crystal structure is affected by crystal contacts between STMV subunits. Crystal contacts were formed entirely by interactions between flexible loops that extend out of the three-fold face, namely loops BC, CD, and GH. Therefore, interactions between these loops from adjacent virions may have influenced the quaternary structure shown in the crystal structure.

4.3.2 The RNA core is stable

Although the capsid arrangement was noticeably different between density maps from crystallography and cryo-EM, the densities corresponding to RNA duplexes were surprisingly similar (Fig 4.1A, B). We fitted the atomic coordinates of the RNA duplexes into the STMV reconstruction without the need of rotational or translational adjustments (Fig. 4.7). Only the five central nucleotides of the nine-base RNA helices could be fitted in the cryo-EM map. These nucleotides make the highest number of contacts with the protein capsid and have the least thermal motion (lowest *B* value), as determined in the crystal structure. We did not observe the inner portion of the RNA helices, which appear disordered in the reconstruction even though full occupancy of the helix has been observed at each dyad in the icosahedrically averaged crystal structure.

These results support previous observations that the RNA core is remarkably stable and that the core has a scaffolding function upon which the capsid proteins are assembled. Thus, the structural role of STMV RNA may be as important as the protein information encoded by the genome. The stability of the RNA structure is consistent with previous reports by atomic force microscopy which indicated that extracted RNA maintains a globular state with a diameter consistent with the inner radius of the capsid shell (*103*). Kuznetsov et. al. reported that rigorous denaturing conditions were needed to disrupt the secondary structure of the RNA (*103*). A molecular dynamic simulation study has also revealed the stability of the RNA core alone (*106*).

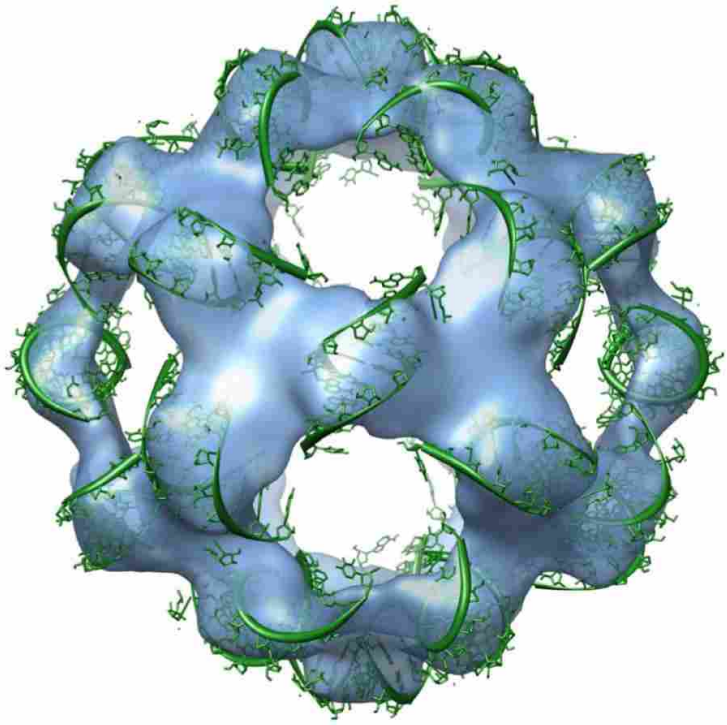


Figure 4.7. STMV RNA coordinates superimposed into the cryo-EM reconstruction. Five central nucleotides out of the nine-base helix were fitted in high contour level densities of the reconstruction.

4.3.3 Capsid expansion at high pH

STMV capsids swell when exposed to high pH (*104*). We investigated STMV capsid expansion by acquiring cryo-EM images of STMV at pH 10. Poliovirus was used as internal standard to calibrate the reconstruction of STMV under this new pH condition. For unknown reasons, poliovirus assumed an 80S conformation (*107*); therefore, we calculated the correct pixel size of the images by comparing poliovirus reconstruction with a previously calibrated 80S poliovirus map. After correcting the pixel size of STMV images, we observed a radial expansion of STMV capsid by ~10Å, double than previously reported (*104*), without a significant change in the overall capsid conformation (Fig. 6.8). Nevertheless, the RNA expands just ~3.5Å in radius. The uncoordinated expansion of the RNA and the capsid suggests that the RNA structure is

inherently more stable than the capsid. We speculate that the RNA as a single molecule has fewer degrees of freedom than the capsid, which is formed by the cooperative arrangement of 60 copies of the capsid protein. The multipartite nature of STMV capsids could contribute to flexibility and expansion of the capsid shell. The discrepancy between the capsid expansion reported here and that previously observed is likely due to methodological differences, the latter being an estimate using atomic force microscopy (*104*).

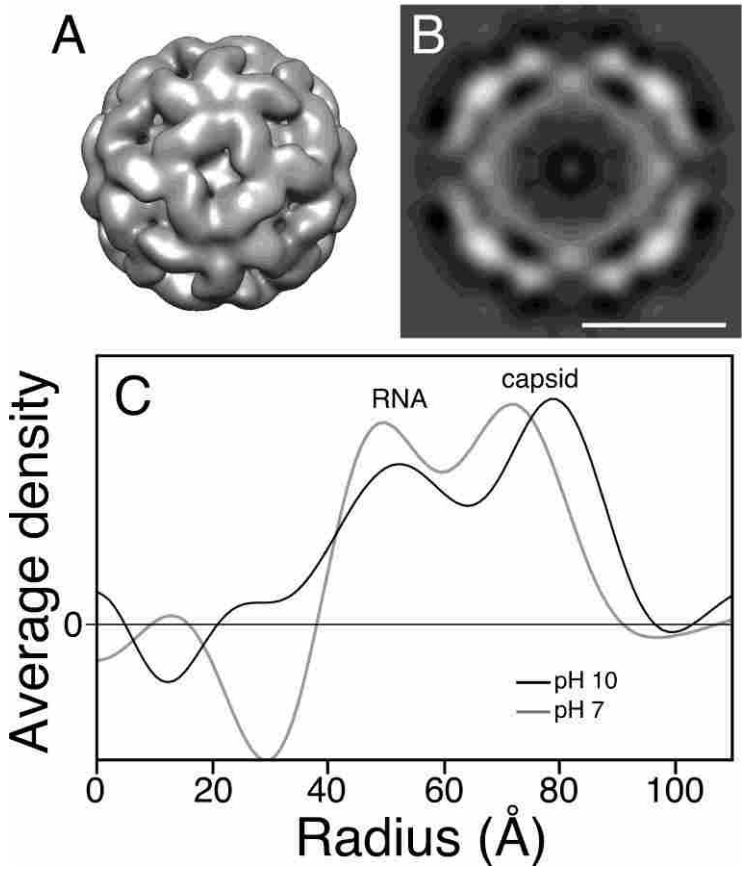


Figure 4.8. STMV reconstruction at pH 10. (A) Surface rendering of the reconstruction at 1-σ contour level. (B) Central slice of the reconstruction. Scale bar, 10 nm. The overall structure is conserved between STMV at pH 7 and 10. (Compare with Fig. 4.1B) (C) Spherically averaged density profiles of STMV cryo-EM structures at pH 7 and 10, showing the expansion of STMV capsid at elevated pH. Conversely, the RNA densities are relatively unchanged.

4.3.4 Structure validation and concluding remarks

To rule out the possibility of model bias in the generation of our reconstruction, we computed ten independent reconstructions using random models, as previously described (*30*). Two out of the ten maps showed identical arrangement of the capsid as the described cryo-EM structure. Only these two maps presented an RNA structure consistent with the crystal structure. Other reconstructions showed uninterpretable structures, such as spiky surfaces or spherical blobs, that are common erroneous products of the random model method (*30, 31*). The relatively poor success rate of the random model method could be attributed to difficulties in orientation determination of STMV particle images. Yet, we recorded different cryo-EM datasets of STMV using various acceleration voltages and defocus ranges, but all reconstructions produced similar structures as our initial results (data not shown).

We speculate that the cryo-EM structure represents a more physiological state of STMV. The current atomic-resolution structure of STMV was derived from crystals formed in 10-16% ammonium sulfate (w/v) (*7*), which is non-physiological. We investigated whether our cryo-EM structure was a result of a favorable conformation due to components in our buffering solution. Yet, cryo-EM of STMV in pure water yielded similar reconstructions as STMV in buffer (data not shown). Our attempts to record cryo-EM images of STMV in a solution containing similar amounts of ammonium sulfate used in crystallography were unsuccessful due to refractory conditions in forming thin, vitreous ice. We postulate that the crystal structure of the virion may represent a natural conformation when STMV is in complex with excess amounts of large anions. Conversely, our cryo-EM structure may represent the STMV structure with physiological concentrations of anions.

The ability to visualize specimens in a vitreous, native-like state is a paramount reason for the growing popularity of cryo-EM. Here, we demonstrated that even low- to modest-resolution structures derived from cryo-EM can be used in complement with crystallography data. Our cryo-EM map, resolved to 14-Å resolution, revealed a significantly different quaternary structure in capsid organization compared to the structure solved by X-ray crystallography. However, the major fold of the capsid protein appears unchanged, as demonstrated by rigid-body fitting of the atomic coordinates into the reconstruction. Removal of all buffer components in the STMV solution did not affect the structure. Therefore, deviations between our cryo-EM structure and the crystal structure are unlikely to have been triggered by the composition of our STMV solution. Yet, capsid variation and dynamics appear to be independent of the RNA structure. Indeed, the ability of STMV capsids to assume various conformations suggests that interactions among the sixty-related subunits are inherently flexible, whereas the RNA core appears to be more rigid. Our results support evidence by others that at least part of the RNA core consistently folds into a stable structure, which presumably drives the assembly process of the virion. The scaffold-like nature of STMV RNA is in line with the model that the RNA core is at least partially folded before capsid proteins tile onto the core. Whether the RNA fully folds into a globular state before assembly or whether assembly occurs in concert with RNA folding is unknown. The consistencies of the RNA structures reported here and by others suggest that the full structure of the STMV genome could be resolved by means of cryo-EM, given sufficient particle images.

STMV particles present a unique challenge for cryo-EM and 3D reconstruction methods. Because of the small size of the virion, ice thickness has to finely tuned. Furthermore, the lack of conspicuous features on the capsid shell could result in incorrect assignment of orientations. Our

laboratory has now obtained monoclonal antibodies to the STMV capsid to use as orientation markers for future cryo-EM experiments. In principle, the antibodies should increase the signal-to-noise ratio of STMV particle images sufficiently to refine and validate our current structures.

4.3.5 Acknowledgments

Co-authors in this study include Eduardo Sanz-García, who is an equal contributor to this work, and David M. Belnap. This work was supported by a grant from the National Institute of Allergy and Infectious Disease (R21AI081221). I thank Alex McPherson and Steven Larson for STMV samples and helpful discussion; Jim Hogle for the poliovirus sample used for size calibration purposes; Jeffrey Farrer for microscopy support; the BYU Fulton Supercomputer Lab for computing resources.

Chapter 5

Structural relationships between halophage CW02 and members of the viral T7 supergroup

5.1 Introduction

Tailed, double-stranded DNA (dsDNA) bacteriophages (order *Caudovirales*) account for ~95% of all documented phages and probably comprise the majority of viruses on the planet. Recent estimations place the global number of virus particles on the order of 10^{31}-10^{32}, outnumbering the total number of host cells by at least 10 to 1 (*1*). The ubiquity of phages suggests an ecological, predatory role in recycling organic matter throughout their environments (*2*). Phages also affect the genetic diversity of their environments via horizontal gene transfer (*3*). The sheer abundance of phages in aquatic environments may cause the greatest selective pressure exerted on host organisms (*2*).

Bacteriophages of the order *Caudovirales* are divided into three morphologically distinct families. All consist of isometric heads that encapsulate linear, dsDNA genomes, but phages of the *Myoviridae* family are characterized by a long, contractile tail (e.g., bacteriophage T4); *Siphoviridae* phages consist of a long, non-contractile tail (e.g., bacteriophage lambda); finally, viruses within the *Podoviridae* family consist of short, non-contractile tails (e.g., bacteriophage T7). Recent high-resolution capsid structures of the tailed, dsDNA phages revealed a conserved, common fold shared among the three phage families. The first of these structures was solved for bacteriophage HK97, and all subsequent structures of *Caudovirales* phages have been shown to

adopt the HK97-like fold. The limited identity between primary structures of HK97-like capsids is a prototypical example that structure is more conserved than sequence.

Here we characterize the structure of CW02, a recently isolated bacteriophage from the Great Salt Lake that infects a strain of *Salinivibrio costicola* bacteria. Based on negative stain transmission electron microscopy, CW02 is morphologically similar to viruses of the *Podoviridae* family, though tail structures were not observed by cryogenic electron microscopy (cryo-EM). Genomic analysis of CW02 revealed strongest sequence similarity with Pseudomonas phage PA11 (*110*), Roseophage SIO1 (*111*), Vibriophage VpV262 (*112*), and Cyanophage Pf-WMP3 (*113*), all of which are distant relatives of bacteriophage T7 but members of the T7 supergroup. The 16-Å resolution structure of CW02 permitted fitting of the HK97-like fold into the cryo-EM map, supporting the observation that many dsDNA phages adopt a common ancestral fold.

5.2 Materials and methods

5.2.1 Cultivation and purification

Host bacteria (*Salinivibrio costicola*) was cultivated in media was supplemented with 8% NaCl. CW02 was initially isolated by plaque assay of filtered water from the Great Salt Lake. CW02 was recovered from plaques by removal from soft agar with a sterile Pasteur pipette. The agar was transferred to saline solution, vortexed, and incubated for 30 minutes to release phage from the matrix. CW02 production was then amplified by inoculation into broth culture of host bacteria. As virus replicates, host cells lyse and the broth culture changes from turbid to clear. After clearing, the broth culture was centrifuged at 5000 × g for 10 minutes and passed through a 0.2 μm filter to remove bacteria.

The crude filtrate was further purified by isopycnic ultracentrifugation (64% CsCl w/v) by spinning at ~125,000 × g for 24 hours at 4°C. A single viral band was extracted from the gradient. CsCl was removed by dialysis in a either pure water or pH 8 buffer containing 1.35M NaCl, 48 mM $MgSO_4 \bullet 7H_2O$, 1 mM $CaCl_2$, and 2 mM Tris-Cl. Finally, the purified CW02 was washed and concentrated by centrifugal filtration (100 kDa molecular weight cutoff filters, Sartorius Stedim, Bohemia, New York, USA) to be made suitable for cryo-EM.

5.2.2 Genome isolation and sequencing

Purified CW02 was concentrated using centrifugal filtration (100kDa cutoff filters, Millipore Corp., Billerica, Massachusetts, USA). Bacterial and contaminant nucleic acids were removed by DNAse and RNAse treatment (Sigma Aldrich Corp., St. Louis, Missouri, USA), then centrifuged at 25,000 × g for 2 hours. CW02 genome was released from capsids by proteinase K and sodium dodecyl sulfate treatment (Sigma). The DNA was then extracted through Phenol-Tris-Cl, followed by a 50/50 mixture of phenol/chloroform, and finally through pure chloroform. The DNA was ethanol precipitated and resuspended in Tris-EDTA buffer. Sequencing was performed via shotgun sequencing by the BYU Sequencing Center using a combination of 454 and Illumina sequencing.

5.2.3 Electron microscopy and image reconstruction

3.5 μl of purified sample was placed on a glow-discharged, holey-carbon-coated copper grid (Electron Microscopy Sciences, Hatfield, Pennsylvania, USA), blotted, and plunge frozen in liquid ethane with an FEI Vitrobot (Hillsboro, Oregon, USA). Specimens were transferred to a Gatan 626 cryoholder (Pleasanton, California, USA) cooled with liquid nitrogen. Cryo-EM images were recorded at 200 keV and at 39,000× magnification via low-dose methods at

objective lens underfocus levels ranging from 0.2–3.8 μm on Kodak SO-163 film in an FEI Tecnai F30 transmission electron microscope (Hillsboro, Oregon, USA).

All electron micrographs were digitized on a Nikon Super Coolscan 9000 ED scanner. Particle images were extracted from electron micrographs using X3DPREPROCESS (*74*) (Table 1). Using Bsoft (*bshow* and *bctf* functions) (*75*), we determined contrast transfer function signal and decay parameters and applied them to correct the images. Origins and orientations of the extracted particles were determined using the model-based technique of *PFT2* (*76*), which was adapted to use phase and amplitude information in orientation selection (*31*). A reference model for *PFT2* analysis was generated by the random-model method with imposed icosahedral symmetry (*30*). The final 3D reconstructions were calculated using *EM3DR2* (*77*, *78*). Contour levels are given in terms of σ, which was calculated as the number of standard deviations relative to the average map density.

For calibration purposes, we used poliovirus (PV, 160S (*107*)) as an internal standard. PV particles were mixed with CW02. The combined sample was prepared for cryo-EM, recorded, and digitized as described above. A previously calibrated map of 160S PV was used to determine the size of the PV reconstruction (*81*) and, hence, the sampling size of the CW02 reconstruction. Bsoft (*bshow* function) (*75*) was used to measure capsid dimensions.

5.2.4 Handedness determination

Tilt experiments were performed as described in (*114*). Briefly, pairs of micrographs were recorded at the same field of view, under the same conditions, but at slightly different tilt angles. The first micrographs of the tilt pairs were recorded at 0° tilt and the second micrographs were recorded at 5° tilt. The origins and orientations for the untilted images were determined as described earlier, after which the orientations of the tilted particles could be predicted. The

predicted orientations were applied to the 3D model to generated tilted projections of either handedness, which were then cross-correlated with the experimentally tilted images to determine the correct handedness. The Bsoft program *dhand* was used to perform the computational tables of the tilt experiment (*75*).

5.2.5 Mass spectrometry

CW02 virions were subject to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to separate individual proteins. The gel was stained using Coomassie blue reagent, after which gel bands were excised then shredded by spinning gel fragments through micropipette tips. Proteins were prepared for analysis by mass spectrometry using a modified filter-aided sample preparation protocol, adapted from (*115*). Briefly, the gel fragments were destained using a 1:1 mixture of acetonitrile:8M urea, 0.1M Tris-HCl (pH 8.5). The denatured proteins were subsequently reduced with 0.1M dithiothreitol in UA buffer (8M urea, 0.1 M Tris-Cl, pH 8.5), then carboamidomethylated with 50 mM iodoacetamide in UA buffer. Each solution was washed by filtration through a 10 kDa molecular weight cutoff filter (Sartorius Stedim, Bohemia, New York, USA). The urea and iodoacetamide were then replaced by a solution of 50mM ammonium bicarbonate. The proteins were finally digested by trypsin and then acidified by addition of formic acid to 1% volume.

The acidified peptide samples were loaded onto a 75 micron x 15 cm nanoAquity C18 column (Waters Corporation, Milford, Massachusetts, USA) and eluted by a 100 min binary gradient of Solvent A (5% acetonitrile) and Solvent B (0.1% formic acid, 99.9% acetonitrile) at 425nL/min composed of these steps: From a baseline at 95% Solvent A, an eight-minute gradient to 90% A was performed, followed by a 65-minute gradient to 65% A, a sevenminute gradient to

5% A for eight minutes, followed by a three-minute return to a 95% A baseline for nine minutes.

The column effluent was directed to a nanoESI spray source on a LTQ-Orbitrap XL (Thermo Scientific, San Jose, California, USA). Data dependent acquisition was performed by coupling a 60,000 resolution survey scan in the Orbitrap with up to the top seven subsequent CID MS/MS scans, acquired in the ion trap (signal threshold of 1000, normalized collision energy of 35%, isolation width of 2 m/z, with a two count dynamic exclusion in a three minute window). A lock mass of a polysiloxane compound was used as an internal mass calibration standard.

MS/MS spectra were converted to Mascot generic format by the *MsConvert* program (http://proteowizard.sourceforge.net/pubs.shtml). Mascot searching was queried against a database constructed from the putative open reading frames (ORFs) of the CW02 genome, which were obtained from ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The peptide search space was expanded by concatenating the CW02 database with a comprehensive database of the *Escherichia coli* ATCC 8739 proteome (GenBank accession no. YP_001723016.1) and a reverse decoy version of the same database. Results were filtered to provide a 1% false discovery rate (FDR) confidence at the peptide level.

5.2.6 Sequence analysis and structural modeling

CW02 protein homologues were determined by PSI-BLAST analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Other protein sequences were obtained from GenBank (see Supplemental Data Section 5.4.2 for accession numbers) and aligned using *ClustalW2* (*82*). Protein secondary structures were predicted using *Psi-pred* (*116*), *COILS* (*117*), *BetaWrapPro* (*118*), or *Phyre* (*119*). Phylogenetic analysis of the structure module proteins (capsid, scaffold,

portal, and terminase proteins) was calculated by parsimony methods using the program *PROTPARS* in the *PHYLIP* package (*120*). The phylogenetic dataset was generated by concatenating the protein sequences of each structure module into a single, continuous sequence.

The atomic coordinates of HK97 capsid protein (PDB accession no. 1OHG (*10*)) were first fitted manually as a rigid body into the electron density map of CW02 using the program *UCSF Chimera* (*32*), after which the fitting was refined by automated fitting using the same program. *UCSF Chimera* was also used to select for the Cα atoms of the core HK97 fold to optimize fitting.

5.3 Results and discussion

5.3.1 Characterization of the CW02 genome

The genome of CW02 consists of linear dsDNA and is 49,391 bp long with a GC content of 47.67%. Its genome encodes at least 44 putative gene products and a single tRNA (Arg; anticodon = AGA) (Table 5.1). Fifteen of the putative open reading frames (ORFs) appear unique to CW02 as PSI-BLAST analysis did not reveal any matches. PSI-BLAST of putative CW02 ORFs resulted in strong matches to *Pseudomonas* phage PA11(*110*), Roseophage SIO1 (*111*), Cyanophage Pf-WMP3 (*113*), and Vibriophage VpV262 (*112*). The genomes of CW02 and PA11 share at least 17 homologous ORFs, though the functions of the PA11 gene products have not yet been examined.

A T7-like DNA polymerase gene (gp57) was identified by PSI-BLAST, suggesting an evolutionary relationship between CW02 and T7 (Table 5.1). However, the genus definition of "T7-like phages" includes the presence of an RNA polymerase gene (*121*), which was not detected in the CW02 genome. The genomes of SIO1 and VpV262 also lack an RNA polymerase

gene (*112*). Recent characterization of the SIO1 and VpV262 genomes concluded that a T7-like superfamily forms an ancestral group that predates the divergence of phages with and without an RNA polymerase gene (*112*).

A key, characteristic feature of the T7 supergroup is the presence of a conserved module of structural genes that function as a single genetic unit (*112*). The structure module of the VpV262-like cluster is arranged such that the genes encoding the terminase (DNA packaging enzyme), portal, scaffold, and capsid proteins are consecutively ordered and separated by few, if any, gaps (Fig. 5.1A). Tail proteins are also a component of the structure module and are generally encoded immediately downstream of the capsid protein (*112*). The structure module of T7 is related to the VpV262-like cluster by inversion, with the portal, scaffold, and capsid proteins sequentially ordered (though the terminase ORF is rearranged). Bioinformatics analysis of the CW02 genome revealed the same module arrangement as in VpV262 (*112*) and Pf-WMP3 (*113*)—a conserved terminase (gp10), portal (gp9), scaffold (gp7), and capsid protein (gp6) closely grouped in sequential order (Table 5.1, Fig. 5.1A). Several large ORFs are found downstream of the capsid proteins, and based on the module arrangement, we speculate these as genes involved in tail assembly (gp5, gp2, gp0, gp79, gp78, and gp74) (Table 5.1). Hence, CW02 likely belongs to the VpV262-like cluster within the T7 supergroup.

The evolutionary relatedness of the CW02 structure module with other members of the T7 supergroup was investigated by phylogenetics analysis. We constructed a phylogenetic tree based on the similarities shared among the structure modules of CW02, PA11, SIO1, Pf-WMP3, VpV262, and T7 (Fig. 5.1B). As expected, CW02 was placed in the same clade as other members of the T7 supergroup. The T7 structure module was identified as the outgroup.

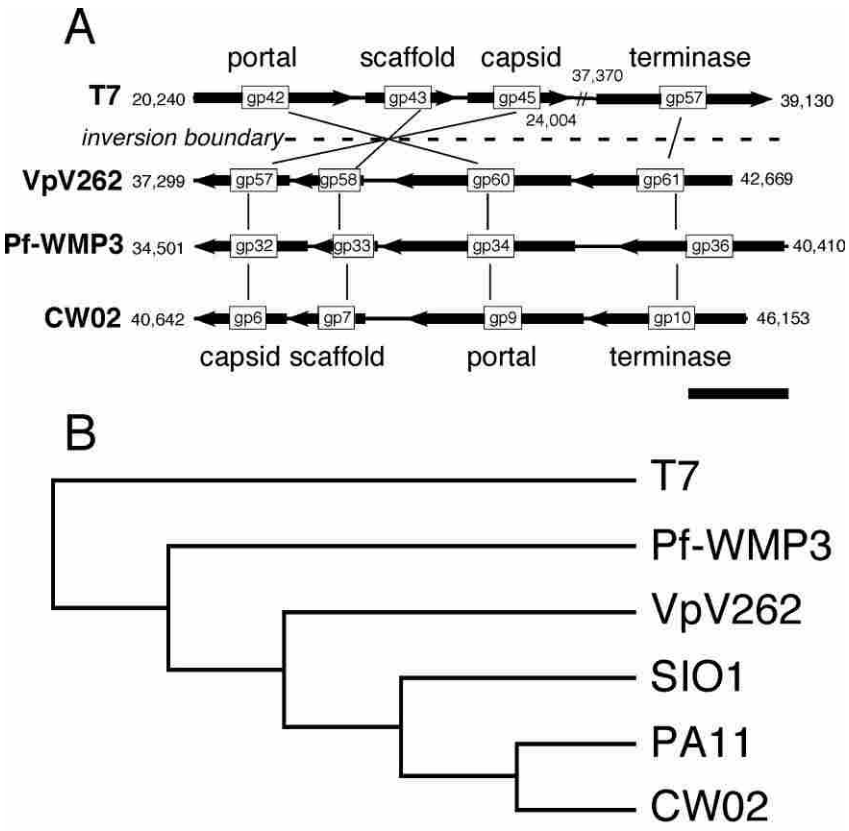


Figure 5.1. The conserved structure module arrangement of the T7 supergroup. (A) Genomic segments of phages T7, VpV262, Pf-WMP3, and CW02. The division of the T7 supergroup (dashed line) is characterized by inverted directionality of the genes in T7 compared to the VpV262-like cluster, as discussed in (*113*). Scale bar, 1kb. (B) Phylogenetic tree constructed from the structure module protein sequences (capsid, scaffold, portal, and terminase) of phages CW02, PA11, SIO1, VpV262, Pf-WMP3, and T7 (outgroup).

Table 5.1. CW02 putative ORFs and analysis of proteins determined by LC-MS and detected homologues.

Gene product	Length[a]	MW (kDa)	LC-MS coverage[b]	BLAST hits[c]	E value	Secondary structure prediction[d]	Function
gp74	608	64.8	49%, 43%	Collagen triple helix repeat protein [Victivallis vadensis ATCC BAA-548] Putative tail fiber [Acinetobacter baumannii 1656-2] Conserved phage tail repeat-like domain	8e-06 6e-05		Possible tail fiber
gp75	102	11.5	-	hypothetical protein [Vibrio phage ICP2]	2e-13		
gp76	118	13.3	-	-			
gp77	144	15.4	-	ORF-B (predicted transmembrane helix, suspected holin) [VpV262]	2e-22	Transme-mbrane helix (*Psipred*)	Suspected holin
gp78	684	73.7	41%, 44%	-		coiled-coils (*COILS*)	
gp79	1413	156.4	42%, 47%	ORF-U [VpV262] gp26 [Roseobacter phage SIO1]	2e-34 2e-13		Putative tail sheath
gp80[e]	153	16.0	-, 19%	-			
gp0	655	71.2	16%, 19%	Hypothetical protein ORF038 [Pseudomonas phage PA11] gp27 [Roseobacter phage SIO1] ORF-P [VpV262]	8e-128 2e-55 7e-24	2 right-handed β-helices (*BetaWrapPro*)	Putative tail appendage
gp2	239	27.5	17%, 27%	Hypothetical protein ORF040 [Pseudomonas phage PA11] gp28 [Roseobacter phage SIO1] ORF-Q [VpV262]	4e-82 2e-21 7e-06		
gp3	115	12.0	-	Head fiber protein [Bacillus phage B103]	8e-03		Possible head fiber
gp5	152	15.4	36%, 25%	-		β -structure (*Psipred*) Right-handed β-helix (*BetaWrapPro*)	Putative tail fiber
gp6	316	34.7	92%, 32%	Hypothetical protein ORF044 [Pseudomonas phage PA11] ORF-K [VpV262] P22 gp6-like protein [Clostridium phage CP26F & 39-O]	6e-109 3e-31 1e-02		Major capsid protein
gp7	264	29.1	-	Hypothetical protein ORF045 [Pseudomonas phage PA11] gp30 [Roseobacter phage SIO1] ORF-J (coiled coil; scaffold protein) [VpV262]	1e-27 6e-09 8e-09	helical structure (*Psipred*) coiled-coil (*COILS*)	Scaffold protein

Gene product	Length[a]	MW (kDa)	LC-MS coverage[b]	BLAST hits[c]	E value	Secondary structure prediction[d]	Function
gp9	592	67.5	33%, 39%	Hypothetical protein ORF047 [Pseudomonas phage PA11] ORJ-H [VpV262] gp3 & gp5 [Roseobacter phage SIO1] gp8 [Phormidium phage Pf-WMP3]	0.0 1e-78 1e-63 4e-09		Portal protein
gp10	535	61.6	-	Hypothetical protein ORF049 [Pseudomonas phage PA11] ORF-G [VpV262] gp1, gp2, & gp4 [Roseobacter phage SIO1] Terminase [Pf-WMP3] Terminase [Enterobacteria phage T7]	0.0 1e-113 1e-97 6e-16 4e-06		DNA packaging enzyme
gp12	636	68.1	53%, 66%	-			
gp13	191	21.1	-	Hypothetical protein ORF052 [Pseudomonas phage PA11]	7e-37		
gp17	210	23.4	-	-			
gp20	116	13.2	-	-			
gp22	158	18.7	-	Conserved hypothetical protein Acear_1218 [Acetohalobium arabaticum DSM 5501]	3e-21		
gp23	188	20.5	-	-			
gp26	125	14.2	-	-			
gp27	121	14.3	-	Cell wall hydrolase, SleB [alpha proteobacterium BAL199] Conserved hydrolase_2 domain	4e-09		Cell wall hydrolase
gp28	144	16.4	-	-			
gp29	152	16.9	-	-			
gp31	503	56.2	-	-			
gp33	215	24.2	-	glutamine amidotransferase [Bacillus sp. B14905]	8e-12		Amidotransfe-rase
gp34	437	48.8	-	glutamine amidotransferase [Enterobacteria phage ϕEco32]	4e-21		Amidotransfe-rase
gp36	267	30.0	-	Conserved hypothetical protein ORF005 [Pseudomonas phage PA11]	1e-68		
gp39	190	21.1	-	Hypothetical protein ORF006 [Pseudomonas phage PA11]	6e-05		
gp40	266	30.7	-	Hypothetical protein ORF007 [Pseudomonas phage PA11] Conserved amidoligase_2 domain	1e-37		amidoligase
gp41	167	18.5	-	-			
gp43	252	28.6	-	Hypothetical protein ORF008 [Pseudomonas phage PA11] ATP-grasp enzyme [Enterobacteria phage phiEco32] Conserved ATP-grasp domain	1e-63 1e-26		ATP-grasp protein

Gene product	Length[a]	MW (kDa)	LC-MS coverage[b]	BLAST hits[c]	E value	Secondary structure prediction[d]	Function
gp47	126	14.0	-	Conserved YtfP/UPF0131 protein [Enterobacteria phage phiEco32] Conserved AIG-2 like domain	8e-25		AIG-2 protein
gp48	111	12.2	-	-			
gp51	553	62.3	-	Hypothetical protein ORF011 [Pseudomonas phage PA11] DNA primase/helicase [Phage T7]	3e-157 7e-24		DNA primase/ helicase
gp55	161	19.3	-	Hypothetical protein ORF012 [Pseudomonas phage PA11] RNA polymerase sigma-70 factor [Kordia algicida OT-1]	8e-13 2e-06		RNA polymerase sigma factor
gp57	629	72.6	-	Hypothetical protein ORF013 [Pseudomonas phage PA11] Putative DNA polymerase [Roseobacter phage SIO1] DNA polymerase [Phage T7]	0.0 2e-22 5e-07		DNA polymerase
gp59	270	29.6	-	Hypothetical protein ORF015 [Pseudomonas phage PA11]	3e-36		
gp60	152	17.5	-	Hypothetical protein ORF018 [Pseudomonas phage PA11] 5'-3' exonuclease [Pseudomonas phage LUZ24]	1e-36 4e-13		5'-3' exonuclease
gp66	129	15.0	-	-			
gp67	256	29.7	-	Hypothetical protein ORF021 [Pseudomonas phage PA11] exonuclease related to epsilon chain of DNA polymerase [Xanthomonas phage Xp10] Conserved 3'-5' exonuclease domain of family-B DNA polymerases	8e-87 3e-34		3'-5' exonuclease related to DNA polymerase, epsilon chain
gp68	115	12.8	-	Transcriptional repressor [Pseudoalteromonas phage H105/1]	3e-16		putative transcriptional repressor
gp72	191	21.6	-	Conserved poly A polymerase head domain			putative poly-A polymerase

[a] amino acid residues
[b] The first number corresponds to coverage detected from excised SDS-PAGE bands of mature virions. The second number corresponds to coverage detected from whole, undenatured virions.
[c] Summary of hits found via PSI-BLAST analysis. Protein source indicated by brackets.
[d] Secondary structure prediction programs indicated in italics
[e] gp80 was not found to migrate to a specific SDS-PAGE position

5.3.2 Mass spectrometry of CW02 proteins

Proteins associated with the mature CW02 virion were analyzed by liquid chromatography-mass spectrometry (LC-MS), following excision of detected SDS-PAGE gel bands. A total of ten bands were excised then treated by in-gel trypsin digestion (Fig. 5.2), from which nine unique proteins were identified (Table 5.1). We also performed LC-MS of whole, non-denatured CW02 virions, which identified the same nine proteins plus an additional protein, gp80, which was not detected via SDS-PAGE. Three of the gel bands were identified as the same protein (gp6) and one of the gel bands revealed two proteins of similar sizes (gp78 and gp0). Three of the proteins identified by LC-MS (gp12, gp78, and gp80) did not have detectable homologous counterparts via PSI-BLAST analysis.

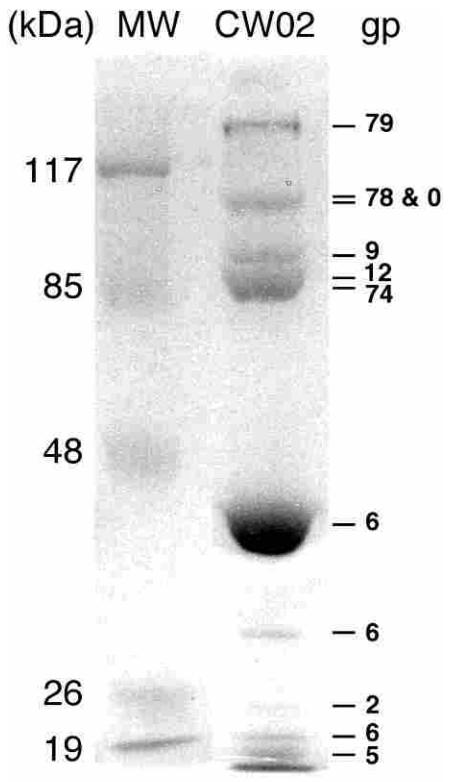


Figure 5.2. Protein composition of mature CW02 particles determined by SDS-PAGE and identified by LC-MS. Gp78 and gp0 have similar sizes and were not separated in the gel.

5.3.3 Analysis of the CW02 structure module

The most abundant CW02 protein identified by SDS-PAGE was gp6, which supported its assignment as the major capsid protein (Fig. 5.2). PSI-BLAST analysis revealed that gp6 shares some similarities with the putative P22-like coat proteins of Clostridium phages CP26F and 39-O (*122*), although the E values were only on the order of 10^{-2} (Table 5.1). PSI-BLAST revealed gp6 to be more homologous with the putative capsid protein of VpV262. We investigated the possible role of gp6 by applying secondary structure prediction methods and comparing the predicted structures with other secondary structure profiles of proteins related to the T7 supergroup. Given that protein structure is generally more conserved than protein sequence, proteins with similar structural profiles may have similar functions. Secondary structure prediction of gp6 showed a similar profile to that of the major capsid protein of phage T7 (Fig. S5.1), even though the sequences are only ~18% identical (55 of 312 residues, Fig. S5.2). We used the Phyre Protein Fold Recognition Server (*119*), which predicted that gp6 adopts the same fold as HK97 and P22 ($p < 0.05$). As observed in the CW02-T7 alignment, pairwise alignments of gp6 with the major capsid proteins of HK97 and P22 revealed just 13% and 20% sequence identity, respectively (Fig. S5.2); multiple gaps were also introduced in the pairwise alignments. The poor similarities between gp6 of CW02 and other established HK97-like proteins are consistent with observations that the HK97-like fold is more conserved than its sequence. As a reference, a pairwise alignment of the HK97 and P22 capsid protein sequences showed just 17% identity (Fig. S5.2).

Based on the arrangement of the CW02 structure module, we speculated that gp7 encodes the CW02 scaffolding protein. PSI-BLAST of gp7 identified similarities with other putative scaffold proteins (Table 5.1). Many viruses require scaffolding proteins to direct the proper

assembly of virus capsids. The proteins are integral components of phage proheads but are not detected in the mature phage particle (*123*). As expected with phage scaffold proteins, we did not detect gp7 by LC-MS of mature virions. We extended our analysis of gp7 by applying secondary structure prediction software, which predicted gp7 to consist of entirely helical structure (Fig. S5.1). The crystal structure of phage ϕ29 scaffolding protein showed that it adopts a predominantly coiled-coil structure (*34*). Although the sequences of gp7 and the scaffolding protein of bacteriophage T7 share only ~15% sequence identity (41 of 264 residues), both proteins are predicted to have similar coiled-coil arrangements (Fig. 5.3, S5.2). Thus, gp7 likely serves as the scaffolding protein of CW02.

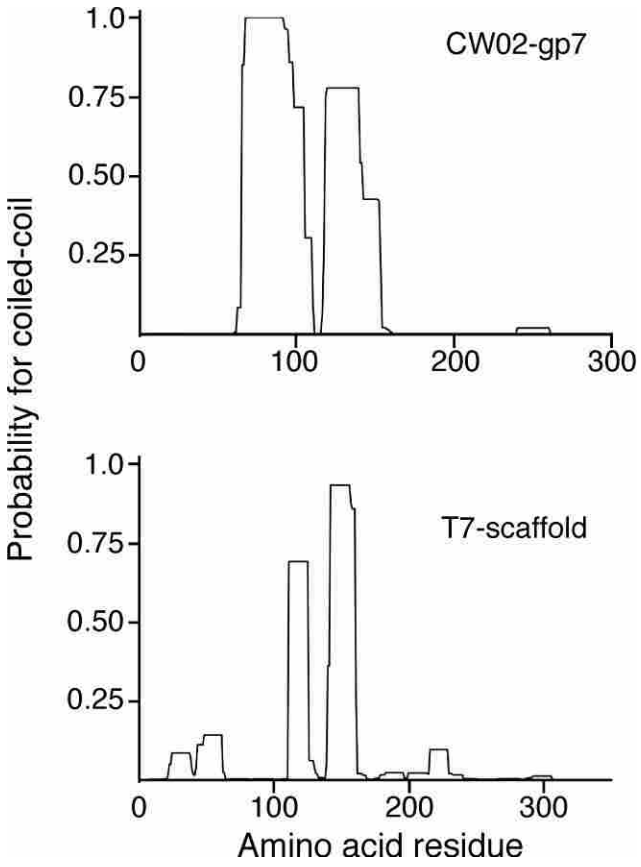


Figure 5.3. Coiled-coil prediction profiles of CW02 gp7 and the scaffold protein of phage T7 (gp9). The probabilities of coiled-coil in either protein were determined using the program *COILS*.

Tailed, dsDNA phages encode portal proteins that enable DNA passage during packaging and ejection, which we speculated to be encoded by CW02 gp9 based on the structure module. The portal also functions as a connector between the capsid and tail proteins. PSI-BLAST of gp9 showed a conserved domain belonging to the head-tail connector superfamily. The sequence of gp9 is strongly conserved with the putative portal proteins of VpV262, SIO1, and *Phormidium* phage Pf-WMP3/4, but not with any established portal proteins (Table 5.1). We used secondary structure prediction methods to investigate possible similarities between gp9 and the portal protein of phage T7 on the structural level and found the profiles to share some resemblance (Fig. S5.1).

Phage terminases are enzymes that facilitate packing of the viral genome into the phage proheads and, like scaffold proteins, are not found in the mature virion. We did not detect the putative terminase protein of CW02, gp10, by LC-MS of mature CW02 particles. The terminase function of gp10 was initially assigned based on sequence homology between gp10 and the terminase protein of phage T7, though the E value was only on the order of 10^{-6} (Table 5.1). The sequence alignments between gp10 and putative portal proteins of VpV262 and SIO1 yielded significantly better E values. Secondary structure prediction of gp10 showed a similar arrangement with the terminase of phage T7, suggesting their functions are conserved (Fig. S5.1).

The structure module of the T7 supergroup places phage tail-encoding genes immediately downstream from the portal, scaffold, and capsids ORF cluster. A number of large ORFs are found in an appropriate location of the CW02 genome that could reasonably be thought to encode tail-associated proteins. For instance, PSI-BLAST analysis of gp74 revealed a conserved "phage tail-like repeat" domain. Also, gp79, the largest putative ORF of the CW02 genome,

encodes a 1,413 amino acid protein—consistent with the large sizes of phage tail sheath proteins. Trimeric structures of coiled-coils and β-helices are a common motif in tail-associated proteins, such as the tail spike protein of P22 (homotrimers of β-helices) (*124*) and tail fibers of bacteriophage T7 (triple-stranded coiled-coils) (*125*). We investigated for the presence of coiled-coils and β-helices in candidate ORFs downstream from the capsid gene. The *BetaWrapPro* program revealed right-handed β-helical motifs in gp0 and gp5. The *COILS* program detected coiled-coil motifs in gp78 and gp80.

5.3.4 Electron microscopy of CW02 particles

Negatively stained CW02 particles were visualized by transmission electron microscopy (Fig. 5.4A). The particles are morphologically similar to members of the family *Podoviridae* based on having an icosahedral head and a short, stubby tail-like feature. We used cryogenic electron microscopy (cryo-EM) to visualize CW02 particles in an unstained, hydrated state (Fig. 5.4B). Interestingly, tail-like structures were not visible in the cryo-EM micrographs.

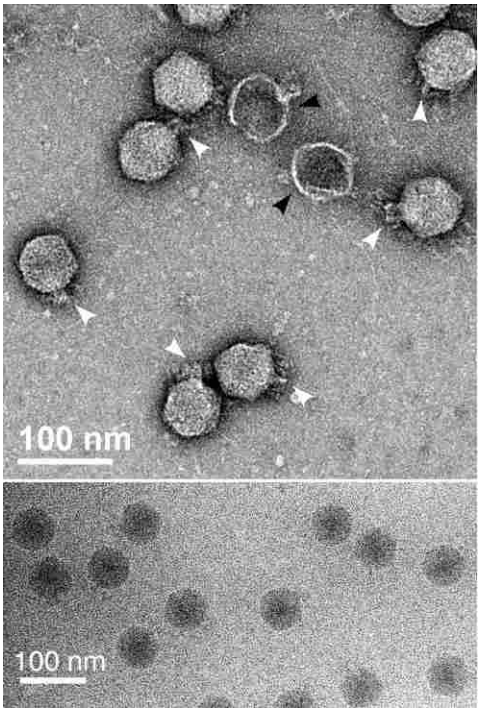


Figure 5.4. Electron microscopy of CW02 by (A) negative stain and (B) cryogenic methods. Stubby, tail-like features are marked by white arrowheads. Collapsed, empty particles shown by black arrowheads. The cryo-EM image shown was recorded at ~2.2 μm underfocus. Tail structures were not detected by cryo-EM.

5.3.5 Architecture of the CW02 head

The three-dimensional, cryo-EM reconstruction of the CW02 virion was determined to 16-Å resolution using 10,677 particle images and applying icosahedral averaging to maximize the signal-to-noise ratio (Fig. 5.5A). The head of CW02 is arranged in a $T = 7$ icosahedral lattice. The diameters of the CW02 head measured 691 Å, 627 Å, and 585 Å along the icosahedral five-fold, three-fold and two-fold axes, respectively. The average thickness of the capsid shell is approximately 37 Å. Sixty hexamers and twelve pentamers comprise the icosahedral head.

To date, the capsids of all tailed, dsDNA bacteriophages appear to adopt the HK97-like fold. For instance, the pseudo-atomic capsid structure of phage T7, solved by cryo-EM and modeling methods, showed the conserved fold (*126, 127*), therefore suggesting CW02 might also

adopt the same fold. The centers between two hexameric capsomeres of the CW02 reconstruction are separated by approximately 140 Å, which is a characteristic feature of the HK97-like fold (*128-130*). The fold consists of two domains: the axial (A) domain and the peripheral (P) domain (*10*) (Fig. 5.6A). The A domain is a compact, wedge-like structure with two α-helices (α5 and α6) and a four-stranded β-sheet. The P domain is elongated, consisting of a long helix (α3) accompanied by a long β-sheet. Together, the domains form an L-shaped protein with a continuous hydrophobic core (*10*). The HK97 capsid protein also bears an extended amino-terminal arm and a cross-linking loop (Fig. 5.6A), but these are not as conserved in other tailed, dsDNA phages.

We performed rigid-body fitting of the HK97 A and P domains into the CW02 reconstruction (Fig. 5.6B-D). The Cα backbone of the A and P domains were fitted as a single entity as either hexamers or pentamers. Flexible loops were omitted from the fitting. Of the 768 Cα atoms used to fit the hexameric capsomere, only 29 (3.7%) protruded from the electron density map. A striking feature of the fitting was the agreement between the characteristic slants of the α5 and α6 helices into the cryo-EM densities (Fig. 5.6B, solid lines). The elongated α3 helix and accompanying P domain β-sheet also fit snugly into the CW02 densities (Fig. 5.6C). A similar result was seen in fitting the pentameric capsomere, with just 15 of 649 Cα atoms (2.3%) protruding from the map. The P domains interact in a threefold fashion at the threefold and quasi-threefold regions. Our fit showed agreement between the P domain coordinates and the CW02 map, though voluminous triangular densities were found above each of the P domain fits (Fig. 5.6B and C, star). The overall fit is consistent with previous studies that showed the intra-capsomere stabilization role of the A domain and the inter-capsomere contacts made by the P domain (*10*).

The HK97 capsid protein structure also consists of a cross-linking loop and extended N-terminal arms that stabilize interactions within each capsomere. No evidence of cross-linking is seen in CW02 due to the absence of gp6 oligomers in SDS-PAGE analysis (Fig. 5.2). The portions of the HK97 cross-linking loop and N-terminal arms fit poorly into the inter-capsomere space, suggesting CW02 has different means of inter-capsomere contacts (data not shown).

The pentameric and hexameric capsomeres appear similar, but the centers of the hexameric capsomeres are curved inwards, whereas the centers of pentameric capsomeres have an exterior bulge (Fig. 5.5). The bulge may correlate with an additional protein complex.

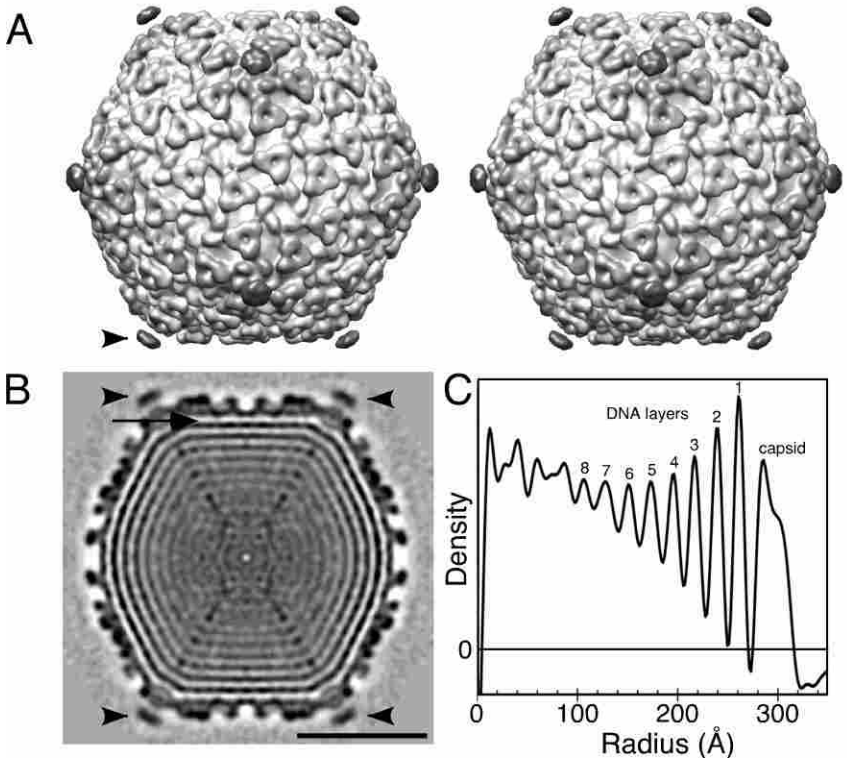


Figure 5.5. Cryo-EM based reconstruction of CW02. (A) Stereo-image of the CW02 capsid surface rendered at 1-σ contour level. (B) The central slice of the reconstruction. Black represents protein or DNA density. Bar represents 25 nm. Black arrowheads indicate turrets. Arrow indicates capsid-dsDNA contacts. (C) Plot of average density with respect to radius of the reconstruction. Alternating DNA densities are labeled.

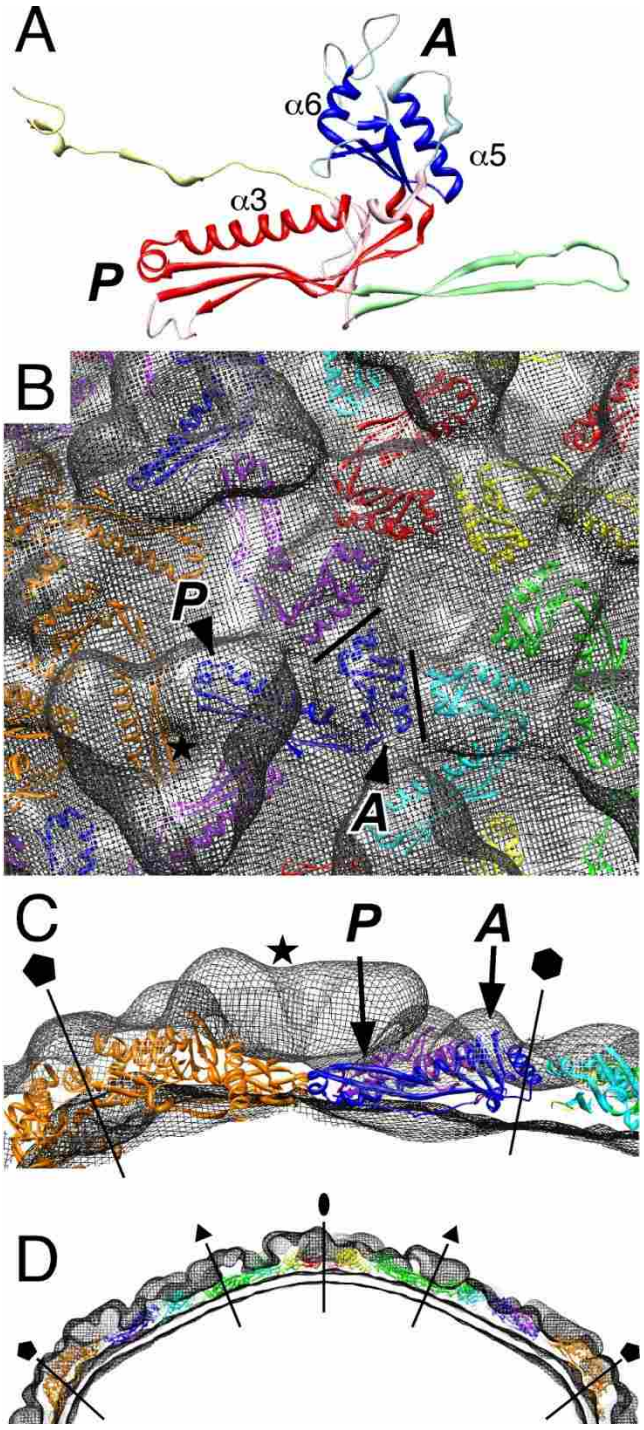


Figure 5.6. Rigid-body fit of the HK97-like fold into CW02 capsomeres. (A) Ribbon diagram of the HK97 protein with pertinent helices and domains labeled. Yellow, N-terminus; green, cross-linking arm; blue, A domain; red, P domain (*10*). Yellow, green, light red, and light blue colors indicate coordinates omitted in the CW02 fitting. (B) Front view of the HK97 A and P domains (chains A-G) fitted into a CW02 hexamer (purple, red, yellow, green, teal, and blue, chains A-F, respectively) and a subunit of a pentamer (orange, chain G). This view emphasizes the A domain. Each subunit is colored according to conventions used in (*10*). Solid lines signify slants of the A domain helices. (C) Side view of the same fit, emphasizing the P domain. Stars in (B) and (C) indicate triangular densities. (A-C) A and P domains indicated in italics. (D) Off-central section of the CW02 reconstruction fitted by multiple copies of the HK97-like fold. Icosahedral components labeled with pentagons (5-fold axis), triangles (3-fold axis), an oval (2-fold axis), and a hexagon (local 6-fold symmetry in panel C).

5.3.6 Handedness of the capsid lattice

T = 7 lattices have two possible arrangements that are mirror images of each other. Because transmission electron microscopy images are two-dimensional projected views, information of the object handedness is lost in its 3D reconstruction. The structures of all currently known T = 7 bacteriophages are of the *laevo* orientation. We performed tilt experiments to determine the handedness of the CW02 capsid. This method relies on cross-correlating images of tilted specimens with projections of the tilted model of either handedness (*114*). The handedness of CW02 was determined to be of the *laevo* orientation, consistent with the structures of other T = 7 bacteriophages, including phages HK97 and T7.

5.3.7 Triangular densities at the threefold and quasi-threefold axes

Each hexamer-hexamer and hexamer-pentamer contact is bridged by a triangular body found on the threefold and quasi-threefold axes (Fig. 5.5 and stars in 5.6B, C). Both types of triangular densities showed consistent structures, despite the icosahedral averaging over the threefold axes, therefore suggesting that the proteins on the threefold and quasi-threefold axes are identical.

The functions of the triangular densities are unknown, but their positions on the threefold and quasi-threefold axes suggest an inter-capsomere stabilization role, similar to the HK97 P-domain. The P-domains of the HK97-like fold form inter-capsomere contacts directly beneath the triangular densities (*10*) (Fig. 5.6). In other HK97-like phages, inter-capsomere interactions by the P-domain are known to be accompanied by more extensive inter-capsomere interactions from additional accessory proteins (as seen in bacteriophages lambda (*131*) and T4 (*132*)), covalent cross-links (as in HK97 (*17*)), or from additional domains as part of the capsid protein. For instance, the pseudoatomic structure of bacteriophage ϕ29 showed that the major capsid protein consists of the HK97-like domain and an additional immunoglobulin-like domain (*34*).

Bacteriophage P22 also consists of extra globular densities on the outermost surface of the capsid (*130*). Unlike the densities seen in CW02, the additional domains in ϕ29 and P22 do not form trimeric structures. Further, the capsid proteins of ϕ29 and P22 are 166 and 147 residues longer than the HK97 protein (448 and 429 vs. 282 residues, respectively), which account for the additional densities that are not part of the HK97-like fold. The size of the CW02 capsid protein (312 residues) is only slightly larger than HK97, suggesting that the triangular densities of CW02 capsids belong to an additional protein complex. If the densities represent separate protein trimers, then 420 copies of the protein would be required per capsid (e.g., the same number of gp6 subunits). Yet, SDS-PAGE analysis of CW02 proteins did not reveal other proteins as abundant as gp6 (Fig. 5.2). Thus, we speculate that the triangular densities associated with CW02 represent an additional domain of gp6, perhaps replacing the cross-linking loop of the HK97 protein (Fig. 5.6A).

5.3.8 Presence of turrets on five-fold vertices

Negative stain and cryo-EM analysis of CW02, respectively, showed either short, stubby tail-like structures or entirely absent tail structures that are normally associated with a unique vertex of the isometric head (Fig. 5.4). No tail-like features were observed in the icosahedrally-averaged structure of CW02 (Fig. 5.5). Instead, each of the icosahedral vertices was marked with the presence of turret-like structures that extends ~40 Å beyond the capsid head. The turret densities appear disconnected from the capsid, suggesting that the linker between turret and capsid is disordered. To investigate whether the turrets are unique to a single vertex or if they are positioned at each of the twelve five-fold vertices of the capsid, we generated reconstructions of CW02 without any symmetric operations. These reconstructions also showed the presence of disconnected turrets on all vertices (data not shown).

Possible explanations for the conflicting data of CW02 tails between negative stain and cryo-EM imaging include the following: negative staining procedures can cause structural distortions in the specimen due to desiccation of the sample and the harsh nature of the uranyl acetate stain. Tail-like densities may also represent material ejected from the capsid during the staining procedure.

Based on genomic analysis of CW02 and its relationship with members of the T7 supergroup, one would expect that each capsid is associated with a short, stubby tail. The fact that none of our reconstructions produced a tail-like structure is consistent with the observation that no tail is visible in the cryo-EM images. The addition of an artificial, cylindrical density to a single vertex of the CW02 capsid as an orientation-refining fiducial marker did not produce different structures (data not shown). The possibility remains that CW02 tails were lost during purification and the tail-like features seen by negative stain analysis represent other entities, but infection of the *Salinivibrio* host with purified CW02 did not result in diminished titer (data not shown). Others have shown that tail-deletion mutants of bacteriophages T7 (*133*) and ϕKZ (*129*) produced stable, DNA-filled capsids, but these are presumably non-infectious.

Our evidence suggests that CW02 lacks a single tail and that the turrets may be unique appendages, similar to those seen in other extremophilic viruses, including *Sulfolobus* turreted icosahedral virus (*15*) and a halophilic virus, SH1 (*134*). Other icosahedral viruses with protein turrets or spikes on vertices include adenovirus (*135*) and bacteriophage PRD1 (*136*). The function of turrets is thought to assist in host adsorption (*134*). To date, capsids with turrets or spikes are shown to adopt a fold consisting of a single or double β-jellyroll fold (see Chapter 1) that is not seen in CW02. Thus, the presence of turrets on CW02 could indicate evolutionary links between viruses of the PRD1-like and HK97-like lineages.

5.3.9 Structure of the encapsidated dsDNA

We observed at least eight concentric layers of densities within the icosahedrally-averaged head of CW02 (Fig. 5.5C, D). Each layer is spaced approximately 24 Å apart, similar to the dsDNA densities observed in other bacteriophage structures, including phages T7 (*133*) and lambda (*137*). The layering pattern is consistent with the DNA spooling model established for T7 (*133*). Weak densities are also observed at the interface between the outermost DNA layer and the capsid (Fig. 5.5C), indicating possible interactions between the capsid protein and dsDNA core.

Some turreted, extremophilic viruses of the PRD1-like lineage, including *Sulfolobus* turreted icosahedral virus and SH1, encase an internal lipid membrane that is thought to aid in injecting the genome into the host cell (*15, 138*). The membrane surrounds the viral genome and undergoes a structural rearrangement into a tail-like structure (*138*). We investigated the presence of internal membranes in CW02, but symmetric and asymmetric reconstructions did not show any evidence of material other than dsDNA within the capsid. The apparent lack of a unique tail and inner membrane suggests that CW02 uses unconventional mechanisms for genome injection. The mechanisms for CW02 infections are unknown, but one possibility may involve a conformational change in the CW02 turrets to form a passageway for genome transfer.

5.4 Appendix

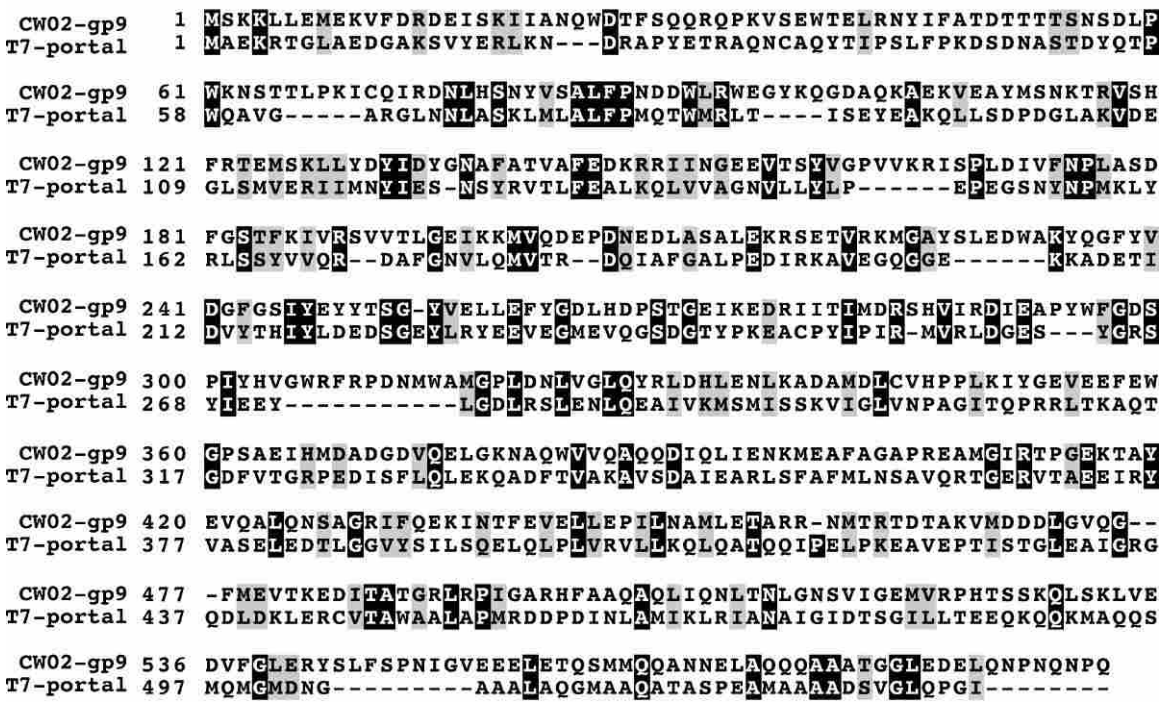
5.4.1 Supplementary Figures

capsid protein alignment

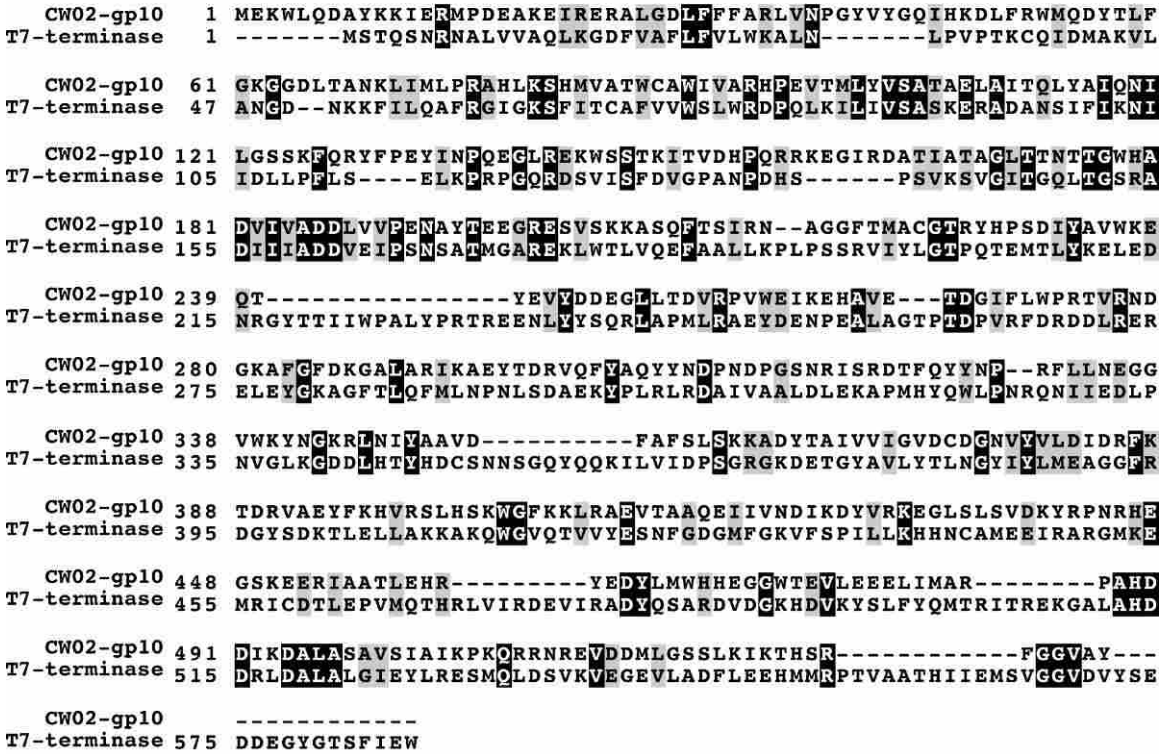
```
 CW02-gp6    1 -------MQLTTNTRAFIEAEQYSSFILTNMHDGLLPQTFYRD--------VSDFGSGTT
T7-capsid    1 MASMTGGQQMGTNQGKGVVAAGDKLALFLKVFGGEVLTAFARTSVTTSRHMVRSISSGKS

 CW02-gp6   46 LHIKTIG--DVSIQDAAEDVALEYSPIDTGEVTLQITDYVGDAWFIVDDLREDGAQVEQL
T7-capsid   61 AQFPVLGRTQAAYLAPGENLDDKRKDIKHTEKVITIDGLL-TADVLIYDIEDAMNHYDVR

 CW02-gp6 104 AAARSAESTRALQERFESRYLEVCEAAQTDADPNK-VNGFAHRIASADTNGKLELSDLIK
T7-capsid 120 SEYTSQLGESLAMAADGAVLAEIAGLCNVESKYNENIEGLGTATVIETTQNKAALTDQVA

 CW02-gp6 163 MK----LAFDKAEVPMAG-RVAIMDPICAAHLDQQVSITNDVTPFAEQILQNGFDREHEF
T7-capsid 180 LGKEIIAALTKARAALTKNYVPAADRVFYCDPDSYSAILAALMPNAANYAAL-IDPEKGS

 CW02-gp6 218 LMNLYGWNIMTSNRLAKG---TFSDGTT------SISDGVANVFMCVADDN-----TKPI
T7-capsid 239 IRNVMGFEVVEVPHLTAGGAGTAREGTTGQKHVFPANKGEGNVKVAKDNVIGLFMHRSAV

 CW02-gp6 264 MAAWRRMPSVEGERNKDLRRDEFVTTSRYGFGPQRKDTLGIYITTAVNI
T7-capsid 299 GTVKLRDLALERARRANFQADQIIAKYAMGHGGLRPEAAGAVVFKVE--


   CW02-gp6   1 MQLTTNTRAFIEAEQYSSFILTNMHDGLLPQTFYRDVSDFGSGTTLHIKTIGDVSIQDAA
HK97-capsid   1 ---SLGSDADSAGSLIQPMQIPGIIMPGLRRLTIRDLLAQGRTSSNALEYVREEVFTNNA

   CW02-gp6  61 EDVALEYSPIDTGEVTLQITDYVGDAWFIVDDLREDGAQVEQLAAARSAESTRALQERFE
HK97-capsid  58 DVVAEKALKPESDITFSKQTANVKTIAHWVQASRQVMDDAPMLQSYINNRLMYGLALKEE

   CW02-gp6 121 SRYLEVCEAAQTDADPNKVNGFAHRIASADTNGKLELSDLIKMKLAFDKAEVPMAGRVAI
HK97-capsid 118 GQLLNGDGTGDNLEGLNKVATAYD--TSLNATGDTRADIIAHAIYQVTESEFSASG--IV

   CW02-gp6 181 MDPICAAHLDQQVSITNDVTPFAEQILQNGFDREHEFLMN-LYGWNIMTSNRLAKGTFSD
HK97-capsid 174 LNPRD----------WHNIALLKDNEGRYIFGGPQAFTSNIMWGLPVVPTKAQAAGTFTV

   CW02-gp6 240 GTTSISDGVANVFMCVADDNTKPIMAAWRRMPSVEGERNKDLRRDEFVTTSRYGFGPQRK
HK97-capsid 224 GGFDMASQVWDRMDATVEVSREDRDNFVKNMLTILCEERLALAHYRPTAIIKGTFSSGS-

   CW02-gp6 300 DTLGIYITTAVNI
HK97-capsid     -------------


  CW02-gp6   1 ------------------------------------MQLTTNTRAFIEAEQYS----
P22-capsid   1 MALNEGQIVTLAVDEIIETISAITPMAQKAKKYTPPAASMQRSSNT-IWMPVEQESPTQE

  CW02-gp6  18 SFILTNMHDGLLPQTFYRDVSDFG------SGTTLHIKTIGDVSIQDAAEDVA------L
P22-capsid  60 GWDLTDKATGLLELNVAVNMGEPDNDFFQLRADDLRDETAYRRRIQSAARKLANNVELKV

  CW02-gp6  66 EYSPIDTGEVTLQITDYVG-----DAWFIVDDLRE---DGAQVEQLAAARSAESTRALQER
P22-capsid 120 ANMAAEMGSLVITSPDAIGTNTADAWNFVADAEEIMFSRELNRDMGTSYFFNPQDYKKAG

  CW02-gp6 119 FESRYLEVCEAAQTDADPN-----KVNGFAHRIAS-------ADTNGKLELS-----DLI
P22-capsid 180 YDLTKRDIFGRIPEEAYRDGTIQRQVAGFDDVLRSPKLPVLTKSTATGITVSGAQSFKPV

  CW02-gp6 162 KMKLAFDKAEVPMAGRVAIMDPICAAHLDQQVSITNDVTPFAEQILQNGFDREHEFLMN-
P22-capsid 240 AWQLDNDGNKVNVDNRFATVTLSATTGMKRGDKISFAGVKFLGQMAKNVLAQDATFSVVR

  CW02-gp6 221 -LYGWNIMTSNR----------LAKGTFSDGTTSISDGVANVFMCVADDNT---------
P22-capsid 300 VVDGTHVEITPKPVALDDVSLSPEQRAYANVNTSLADAMAVNILNVKDARTNVFWADDAI

  CW02-gp6 261 ----KPIMAAWRRMPSVEGE--RNKDLRRDEFVTTS--------------RYGFGPQRKD
P22-capsid 360 RIVSQPIPANHELFAGMKTTSFSIPDVGLNGIFATQGDISTLSGLCRIALWYGVNATRPE

  CW02-gp6 301 TLGIYITTAVNI
P22-capsid 420 AIGVGLPGQTA-
```

capsid protein alignment (cont'd)

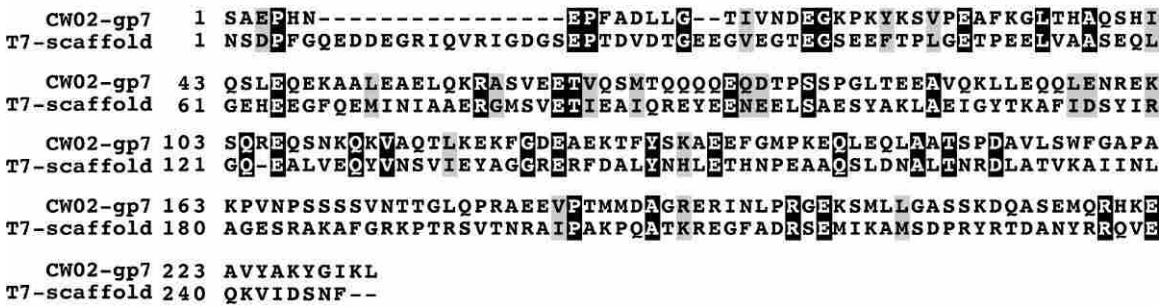
portal protein alignment



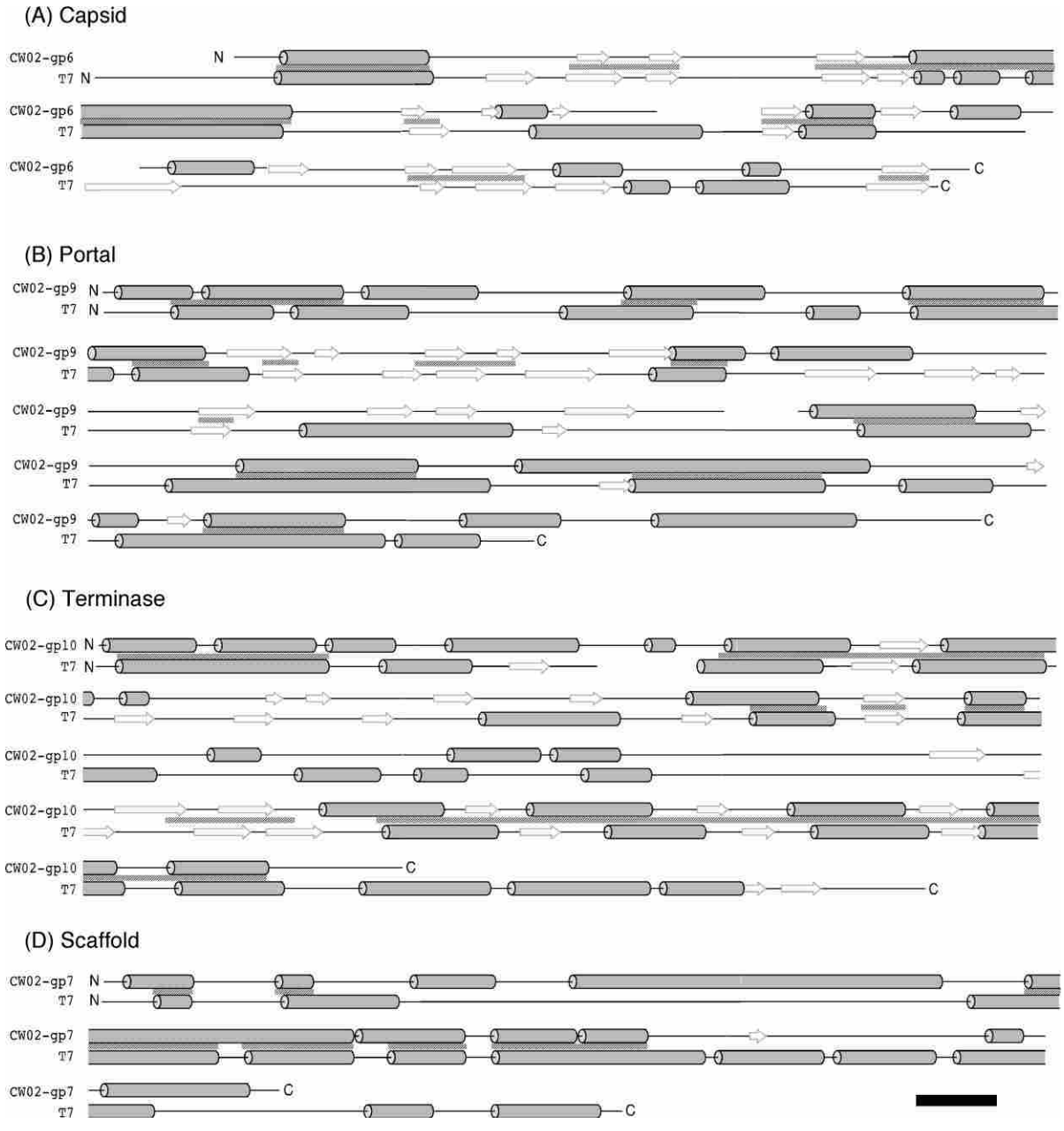
terminase protein alignment



scaffold protein alignment



Supplementary Figure 5.1. (Fig. S5.1) Pairwise sequence alignments of CW02 proteins with various structural phage proteins. Alignments between HK97 and P22 capsid proteins are shown as a reference for the poorly conserved sequences. For the most part, the sequences are similarly poorly conserved between CW02 and other proteins. Black boxes show identical residues. Gray boxes show amino acids with conserved properties.



Supplementary Figure 5.2. (Fig. S5.2) Predicted secondary structures of structure module proteins in CW02 and T7, including (A) capsid, (B) portal, (C) terminase, and (D) scaffold. Similarities in predicted secondary structures between non-conserved sequences may be indicative of conserved structure and function. Cylinders denote helices; arrows denote strands. Scale bar: 10 amino acid residues. Regions of similarity marked by hashed lines in between the alignments. Gaps (breaks in lines) were added subjectively to match primary and predicted secondary structures.

5.4.2 Protein sequence accession numbers

All sequences were obtained from Genbank (http://www.ncbi.nlm.nih.gov/genbank/)

Bacteriophage T7: NC_001604

VpV262: NC_003907

SIO1: NC_002519

Pf-WMP3: NC_009551

PA11: NC_007808

5.4.3 Acknowledgments

Co-authors in this study include Matthew Domek and David M. Belnap. This research was supported by a BYU Graduate Research Fellowship, a Roland K. Robins Graduate Research Fellowship, and a Jennie R. Swensen Graduate Research Fellowship. The BYU Fulton Supercomputing Lab provided computer resources.

I thank Don Breakwell and the BYU Sequencing Center for help with genomic sequencing; Ryan Taylor, Aman Makaju, and John Prince for assistance in mass spectrometry and data interpretation; Robert Swenson and David Eng for help with image processing; Kristin Pande for laboratory assistance; Jeffrey Farrer and Michael Standing for microscopy support; Eduardo Sanz-García for insightful discussions and help with computer programs.

Bibliography

1. K. E. Wommack, R. R. Colwell, Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69 (2000).
2. C. A. Suttle, Marine viruses - major players in the global ecosystem. *Nat. Rev. Micrbiol.* **5**, 801 (2007).
3. L. Van Valen, A new evolutionary law. *Evol. Theor.* **1**, 1 (1973).
4. R. Belshaw *et al.*, Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci U S A* **101**, 4894 (2004).
5. P. S. Moore, Y. Chang, Why do viruses cause cancer? Highlights of the first century of human tumour virology. *Nat. Rev. Cancer* **10**, 878 (2010).
6. F. Crick, J. D. Watson, Virus Structure: General Principles. *Ciba Foundation Symposium on the Nature of Viruses.* Little, Brown, and Company, **5** (1957).
7. S. Larson, J. Day, A. Greenwood, A. McPherson, Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution. *J. Mol. Biol.* **277**, 37 (1998).
8. D. L. Caspar, A. Klug, Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1 (1962).
9. T. S. Baker, J. Drak, M. Bina, The Capsid of Small Papova Viruses Contains 72 Pentameric Capsomeres: Direct Evidence from Cryo-Electron-Microscopy of Simian Virus 40. *Biophys. J.* **55**, 243 (1989).
10. C. Helgstrand *et al.*, The Refined Structure of a Protein Catenane: The HK97 Bacteriphage Capsid at 3.44 Å Resolution. *J. Mol. Biol.* **334**, 885 (2003).
11. M. G. Rossmann *et al.*, Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**, 145 (1985).

12. J. M. Hogle, M. Chow, D. J. Filman, Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* **229**, 1358 (1985).

13. N. G. Abrescia *et al.*, Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**, 68 (2004).

14. S. D. Benson, J. K. H. Bamford, D. H. Bamford, R. M. Burnett, Does common architecture reveal a viral lineage spanning all three domains of life? *Mol. Cell* **16**, 673 (2004).

15. R. Khayat *et al.*, Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc Natl Acad Sci U S A* **102**, 18944 (2005).

16. D. H. Bamford, J. M. Grimes, D. I. Stuart, What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* **15**, 655 (2005).

17. W. R. Wikoff *et al.*, Topologically Linked Protein Rings in the Bacteriophage HK97 Capsid. *Science* **289**, 2129 (2000).

18. M. L. Baker, W. Jiang, F. Rixon, W. Chiu, Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol.* **79**, 14967 (2005).

19. T. Dalianis, R. L. Garcea, Welcome to the *Polyomaviridae*. *Sem. Cancer Biol.* **10**, 209 (2009).

20. D. N. Phalen, V. G. Wilson, D. L. Graham, Organ distribution of avian polyomavirus DNA and virus-neutralizing antibody titers in healthy adult budgerigars. *Am. J. Vet. Res.* **54**, 2040 (1993).

21. P. S. Shen *et al.*, The structure of avian polyomavirus reveals variably sized capsids, non-conserved inter-capsomere interactions, and a possible location of the minor capsid protein VP4. *Virology* **411**, 142 (2011).

22. Y. G. Kuznetsov, J. J. Dowell, J. A. Gavira, J. D. Ng, A. McPherson, Biophysical and atomic force microscopy characterization of the RNA from satellite tobacco mosaic virus. *Nuc. Acid. Res.* **38**, 8284 (2010).

23. R. Johne *et al.*, Avian polyomavirus mutants with deletions in the VP4-encoding region show deficiencies in capsid assembly and virus release, and have reduced infectivity in chicken. *J. Gen. Virol.* **88**, 823 (2007).

24. T. Zhou *et al.*, Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**, 732 (2007).

25. B. D. Welch *et al.*, Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. *J. Virol.* **84**, 11234 (2010).

26. J. Dubochet, M. Adrian, J. Lepault, A. W. McDowall, Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. *Trends Biochem. Sci.* **10**, 143 (1985).

27. R. Grassucci, D. Taylor, J. Frank, Preparation of macromolecular complexes for cryo-electron microscopy. *Nat. Protoc.* **2**, 3239 (2007).

28. N. Grigorieff, S. C. Harrison, Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. *Curr. Opin. Struct. Biol.* **21**, 265 (2011).

29. A. C. Steven, D. M. Belnap, Electron Microscopy and Image Processing: An Essential Tool for Structural Analysis of Macromolecules. *Curr. Prot. Protein Sci.* **17**, 1 (2005).

30. X. Yan, K. A. Dryden, J. Tang, T. S. Baker, Ab initio random model method facilitates 3D reconstruction of icosahedral particles. *J. Struct. Biol.* **157**, 211 (2007).

31. E. Sanz-García, A. B. Stewart, D. M. Belnap, The random-model method enables *ab initio* three-dimensional reconstruction of asymmetric particles and determination of particle symmetry. *J. Struct. Biol.* **171**, 216 (2010).

32. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605 (2004).

33. L. G. Trabuco, E. Villa, K. Mitra, J. Frank, K. Schulten, Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673 (2008).

34. M. C. Morais *et al.*, Conservation of the Capsid Structure in Tailed dsDNA Bacteriophages: the Pseudoatomic Structure of φ29. *Mol. Cell* **18**, 149 (2005).

35. Z. H. Zhou, Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr. Opin. Struct. Biol.* **18**, 218 (2008).

36. X. Yu, L. Jin, Z. H. Zhou, 3.88 Å structure of cytoplasmic polyhedrosis virus by cryoelectron microscopy. *Nature* **453**, 415 (2008).

37. M. A. Marti-Renom *et al.*, Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291 (2000).

38. K. Arnold, L. Bordoli, J. Kopp, T. Schwede, The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195 (2006).

39. B. Rost, Twilight zone of protein sequence alignments. *Protein Eng.* **12**, 85 (1999).

40. J. T. Finch, The Surface Structure of Polyoma Virus. *J. Gen. Virol.* **24**, 359 (1974).

41. I. Rayment, T. S. Baker, D. L. D. Caspar, W. T. Murakami, Polyoma Virus Capsid Structure at 22.5 Å Resolution. *Nature* **295**, 110 (1982).

42. D. M. Salunke, D. L. D. Caspar, R. L. Garcea, Self-Assembly of Purified Polyomavirus Capsid Protein VP1. *Cell* **46**, 895 (1986).

43. T. Stehle, Y. Yan, T. L. Benjamin, S. C. Harrison, Structure of Murine Polyomavirus Complexed with an Oligosaccharide Receptor Fragment. *Nature* **369**, 160 (1994).

44. T. Stehle, S. J. Gamblin, Y. Yan, S. C. Harrison, The structure of simian virus 40 refined at 3.1 å resolution. *Structure* **4**, 165 (1996).

45. T. S. Baker, J. Drak, M. Bina, Reconstruction of the Three-Dimensional Structure of Simian Virus 40 and Visualization of the Chromatin Core. *Proc. Natl. Acad. Sci. USA* **85**, 422 (1988).

46. R. C. Liddington *et al.*, Structure of simian virus 40 at 3.8-Å resolution. *Nature* **354**, 278 (1991).

47. D. M. Belnap *et al.*, Conserved Features in Papillomavirus and Polyomavirus Capsids. *J. Mol. Biol.* **259**, 249 (1996).

48. M. Kawano, Xing, L., Tsukamoto, H., Inoue, T., Handa, H., Cheng, R. H., Calcium bridge triggers capsid disassembly in the cell entry process of simian virus 40. *J. Biol. Chem.* **284**, 34703 (2009).

49. J. Nilsson *et al.*, Structure and Assembly of a T = 1 Virus-Like Particle in BK Polyomavirus. *J. Virol.* **79**, 5337 (2005).

50. T. Stehle, S. C. Harrison, Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. *Structure* **4**, 183 (1996).

51. N. Yokoyama, Kawano, M., Tsukamoto, H., Enomoto, T., Inoue, T., Takahashi, R., Nakanishi, A., Imai, T., Wada, T., Handa, H., Mutational analysis of the carboxyl-terminal region of the SV40 major capsid protein VPp1. *J. Biochem.* **141**, 279 (2007).

52. W.-C. Ou, Chen, L.-H., Wang, M., Hseu, T.-H., Chang, D., Analysis of minimal sequences on JC virus VP1 required for capsid assembly. *J. Neurovirol.* **7**, 298 (2001).

53. J. P. Griffith, D. L. Griffith, I. Rayment, W. T. Murakami, D. L. D. Caspar, Inside polyomavirus at 25-Å resolution. *Nature* **355**, 652 (1992).

54. X. S. Chen, T. Stehle, S. C. Harrison, Interaction of polyomavirus internal protein VP2 with the major capsid protein VP1 and implications for participation of VP2 in viral entry. *EMBO J.* **17**, 3233 (1998).

55. R. Johne, H. Müller, Polyomaviruses of birds: etiologic agents of inflammatory diseases in a tumor virus family. *J. Virol.* **81**, 11554 (2007).

56. A. S. Arroube, M. Y. Halami, R. Johne, G. M. Dorrestein, Mortality due to polyomavirus infection in two nightjars (Caprimulgus europaeus). *J. Avian Med. Surg.* **23**, 136 (2009).

57. S. L. Lafferty, A. M. Fudge, R. E. Schmidt, V. G. Wilson, D. N. Phalen, Avian polyomavirus infection and disease in a green aracaris (*Pteroglossus viridus*). *Avian Dis.* **43**, 577 (1999).

58. K. A. Crandall, M. Pérez-Losada, R. G. Christensen, D. A. McClellan, R. P. Viscidi, Phylogenomics and molecular evolution of polyomaviruses. *Adv. Exp. Med. Biol.* **577**, 46 (2006).

59. M. Pérez-Losada *et al.*, Comparing phylogenetic codivergence between polyomaviruses and their hosts. *J. Virol.* **80**, 5663 (2006).

60. R. Stoll, D. Luo, B. Kouwenhoven, G. Hobom, H. Müller, Molecular and biological characteristics of avian polyomaviruses: isolates from different species of birds indicate that avian polyomaviruses form a distinct subgenus within the polyomavirus genus. *J. Gen. Virol.* **74**, 229 (1993).

61. R. Johne, H. Müller, Avian polyomavirus agnoprotein 1a is incorporated into the virus particle as a fourth structural protein, VP4. *J. Gen. Virol.* **82**, 909 (2001).

62. R. Daniels, D. Sadowicz, D. N. Hebert, A *very* late viral protein triggers the lytic release of SV40. *PLoS Pathogens* **3**, e98 (2007).

63. R. Johne, A. Jungmann, H. Müller, Agnoprotein 1a and agnoprotein 1b of avian polyomavirus are apoptotic inducers. *J. Gen. Virol.* **81**, 1183 (2000).

64. R. Johne, H. Müller, Nuclear localization of avian polyomavirus structural protein VP1 is a prerequisite for the formation of virus-like particles. *J. Virol.* **78**, 930 (2004).

65. W. H. Landschulz, P. F. Johnson, S. L. McKnight, The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759 (1988).

66. H. a. N. Müller, R., A polyoma-like virus associated with an acute disease of fledgling budgerigars (*Melopsittacus undulatus*). *Med. Microbiol. Immunol.* **175**, 1 (1986).

67. K. Shiraki, M. Kudou, S. Fujiwara, T. Imanaka, M. Takagi, Biophysical Effect of Amino Acids on the Prevention of Protein Aggregation. *J. Biochem* **132**, 591 (2002).

68. D. A. Vacante, Traub, R., and Major, E.O., Extension of JC virus host range to monkey cells by insertion of a simian virus 40 enhancer into the JC virus regulatory region. *Virology* **170**, 353 (1989).

69. E. O. Major, Miller, A.E., Mourrain, P., Traub, R.G., de Widt, E., Sever, J., Establishment of a line of human fetal glial cells that supports JC virus multiplication. *Proc Natl Acad Sci U S A* **82**, 1257 (1985).

70. M. T. Pho, Ashok, A., and Atwood, W.J., JC virus enters human glial cells by clathrin-dependent receptor-mediated endocytosis. *J. Virol.* **74**, 2288 (2000).

71. M. Jiang, Abend, J.R., Tsai, B., and Imperiale, M.J., Early events during BK virus entry and disassembly. *J. Virol.* **83**, 1350 (2009).

72. L. Pelkmans, Kartenbeck, J., and Helenius, A., Caveolar endocytosis of simian virus 40 reveals a new two-step vesicular-transport pathway to the ER. *Nat. Cell Biol.* **3**, 473 (2001).

73. J. N. Brady, V. D. Winston, R. A. Consigli, Dissociation of polyoma virus by the chelation of calcium ions found associated with purified virions. *J. Virol.* **23**, 717 (1977).

74. J. F. Conway, A. C. Steven, Methods for reconstructing density maps of "single" particles from cryoelectron micrographs to subnanometer resolution. *J. Struct. Biol.* **128**, 106 (1999).

75. J. B. Heymann, D. M. Belnap, Bsoft: Image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3 (2007).

76. T. S. Baker, and Cheng, R. H., A model-based approach for determining orientations of biological macromolecules imaged by cryo-electron microscopy. *J. Struct. Biol.* **116**, 120 (1996).

77. R. A. Crowther, Amos, L. A., Finch, J. T., DeRosier, D. J., Klug, A., Three Dimensional Reconstructions of Spherical Viruses by Fourier Synthesis from Electron Micrographs. *Nature* **226**, 421 (1970).

78. S. D. Fuller, Butcher, S. J., Cheng, R. H., Baker, T. S., Three-dimensional reconstruction of icosahedral particles--the uncommon line. *J. Struct. Biol.* **116**, 48 (1996).

79. H. Gao, M. Valle, M. Ehrenberg, J. Frank, Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *J. Struct. Biol.* **147**, 283 (2004).

80. E. J. Mancini, M. Clarke, B. E. Gowen, T. Rutten, S. D. Fuller, Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. *Mol. Cell* **5**, 255 (2000).

81. D. M. Belnap, W. D. Grochulski, N. H. Olson, T. S. Baker, Use of radial density plots to calibrate image magnification for frozen-hydrated specimens. *Ultramicroscopy* **48**, 347 (1993).

82. M. A. Larkin *et al.*, ClustalW and ClustalX version 2. *Bioinformatics* **23**, 2947 (2007).

83. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, PROCHECK- a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283 (1993).

84. X. Wu, Milne, J. L. S., Borgnia, M. J., Rostapshov, A. V., Subramaniam, S., Brooks, B. R., A core-weighted fitting method for docking atomic structures into low-resolution maps: application to cryo-electron microscopy. *J. Struct. Biol.* **141**, 63 (2003).

85. M. Carrillo-Tripp *et al.*, VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nuc. Acid. Res.* **37**, D436 (2009).

86. J. E. J. Mills, P. M. Dean, Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput-Aided. Mol. Des.* **10**, 607 (1996).

87. M. van Heel, Similarity measures between images. *Ultramicroscopy* **21**, 95 (1987).

88. U. Neu, K. Woellner, G. Gauglitz, T. Stehle, Structural basis of GM1 ganglioside recognition by simian virus 40. *Proc. Natl. Acad. Sci. USA* **105**, 5219 (2008).

89. T. Stehle, S. C. Harrison, High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. *EMBO J.* **16**, 5139 (1997).

90. G. Christiansen, T. Landers, J. Griffith, P. Berg, Characterization of components released by alkali disruption of simian virus 40. *J. Virol.* **21**, 1079 (1977).

91. R. B. Davis *et al.*, A viral disease of fledgling budgerigars. *Avian Diseases* **25**, 179 (1981).

92. G. Bernier, M. Morin, G. Marsolais, A generalized inclusion body disease in the budgerigar (*Melopsittacus undulatus*) caused by a papovavirus-like agent. *Avian Diseases* **25**, 1083 (1981).

93. C. J. Randall, S. Lees, D. M. Inglis, Papovavirus-like infection in budgerigars. *Avian Pathol.* **16**, 623 (1987).

94. M. J. Dykstra, L. H. Bozeman, A light and electron microscopic examination of budgerigar fledgling disease virus in tissue and in cell culture. *Avian Pathol.* **11**, 11 (1982).

95. R. E. D. Rodgers, D. Chang, X. Cai, R. A. Consigli, Purification of recombinant budgerigar fledgling disease virus VP1 capsid protein and its ability for in vitro capsid assembly. *J. Virol.* **68**, 3386 (1994).

96. M. Wolf, R. L. Garcea, N. Grigorieff, S. C. Harrison, Subunit interactions in bovine papillomavirus. *Proc. Natl. Acad. Sci. USA* **107**, 6298 (2010).

97. D. Willits *et al.*, Effects of the Cowpea chlorotic mottle bromovirus β-hexamer structure on virion assembly. *Virology* **306**, 280 (2003).

98. A. Pappachan, C. Subashchandrabose, P. S. Satheshkumar, H. S. Savithri, M. R. N. Murthy, Structure of recombinant capsids formed by the β-annulus deletion mutant — rCP (Δ48–59) of Sesbania mosaic virus. *Virology* **375**, 190 (2008).

99. K. An, S. A. Smiley, E. T. Gillock, W. M. Reeves, R. A. Consigli, Avian polyomavirus major capsid protein VP1 interacts with the minor capsid proteins and is transported into the cell nucleus but does not assemble into capsid-like particles when expressed in the baculovirus system. *Virus Res.* **64**, 173 (1999).

100. K. An, E. T. Gillock, J. A. Sweat, W. M. Reeves, R. A. Consigli, Use of the baculovirus system to assemble polyomavirus capsid-like particles with different polyomavirus structural proteins: analysis of the recombinant assembled capsid-like particles. *J. Gen. Virol.* **80**, 1009 (1999).

101. R. A. Valverde, J. A. Dodds, Evidence for a satellite RNA associated naturally with the U5 strain and experimentally with the U1 strain of tobacco mosaic virus. *J. Gen. Virol.* **67**, 1875 (1986).

102. J. A. Dodds, Satellite tobacco mosaic virus. *Ann. Rev. Phytopath.* **36**, 295 (1998).

103. Y. G. Kuznetsov, S. Daijogo, J. Zhou, B. L. Semler, A. McPherson, Atomic Force Microscopy Analysis of Icosahedral Virus RNA. *J. Mol. Biol.* **347**, 41 (2005).

104. Y. G. Kuznetsov, S. Larson, J. Day, A. Greenwood, A. McPherson, Structural transitions of satellite tobacco mosaic virus particles. *Virology* **284**, 223 (2001).

105. S. B. Larson, A. McPherson, Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Curr. Opin. Struct. Biol.* **11**, 59 (2001).

106. P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, K. Schulten, Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* **14**, 437 (2006).

107. D. M. Belnap *et al.*, Molecular tectonic model of virus structural transitions: the putative cell entry states of poliovirus. *J. Virol.* **74**, 1342 (2000).

108. W. Wriggers, Using Situs for the integration of multi-resolution structures. *Biophys. Rev.* **2**, 21 (2010).

109. A. Fiser, R. Do, A. Sali, Modeling of loops in protein structures. *Protein Sci.* **9**, 1753 (2000).

110. T. Kwan, J. Liu, M. DuBow, P. Gros, J. Pelletier, Comparative Genomic Analysis of 18 *Pseudomonas aeruginosa* Bacteriophages. *J. Bacteriol.* **188**, 1184 (2006).

111. F. Rohwer *et al.*, The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. *Limnol. Oceanog.* **45**, 408 (2000).

112. S. C. Hardies, A. M. Comeau, P. Serwer, C. A. Suttle, The complete sequence of marine bacteriophage VpV262 infecting *Vibrio parahaemolyticus* indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment. *Virology* **310**, 359 (2003).

113. X. Liu *et al.*, Genomic Analysis of Freshwater Cyanophage Pf-WMP3 Infecting Cyanobacterium *Phormidium foveolarum*: The Conserved Elements for a Phage. *Microb. Ecol.* **56**, 671 (2008).

114. D. M. Belnap, N. H. Olson, T. S. Baker, A method for establishing the handedness of biological macromolecules. *J. Struct. Biol.* **120**, 44 (1997).

115. J. R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359 (2009).

116. K. Bryson *et al.*, Protein structure prediction servers at University College London. *Nuc. Acid. Res.* **33**, W36 (2005).

117. A. Lupas, M. Van Dyke, J. Stock, Predicting Coiled Coils from Protein Sequences. *Science* **252**, 1162 (1991).

118. A. V. McDonnell *et al.*, Prediction and comparative modeling of sequences directing beta-sheet proteins by profile wrapping. *Proteins: Struct., Funct., Bioinform.* **63**, (2006).

119. L. A. Kelley, M. J. E. Sternberg, Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363 (2009).

120. J. Felsenstein, PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164 (1989).

121. M. H. V. van Regenmortel *et al.*, Virus Taxonomy: The Classification and Nomenclature of Viruses. *VIIth Report of the ICTV. Academic Press, San Diego*, (2000).

122. B. S. Seal *et al.*, *Clostridium perfringens* bacteriophages φCP39O and φCP26F: genomic organization and proteomic analysis of the virions. *Arch. Virol.* **156**, 25 (2011).

123. R. W. Hendrix, S. R. Casjens, Assembly of bacteriophage lambda heads: protein processing and its genetic control in petit lambda assembly. *J. Mol. Biol.* **91**, 187 (1975).

124. S. Steinbacher *et al.*, Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. *O*-antigen receptors. *Proc. Natl. Acad. Sci. USA* **93**, 10584 (1996).

125. A. C. Steven *et al.*, Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. *J. Mol. Biol.* **200**, 351 (1988).

126. X. Agirrezabala *et al.*, Quasi-Atomic Model of Bacteriophage T7 Procapsid Shell: Insights into the Structure and Evolution of a Basic Fold. *Structure* **15**, 461 (2007).

127. A. Ionel *et al.*, Molecular rearrangements involved in the capsid shell maturation of bacteriophage T7. *J. Biol. Chem.* **286**, 234 (2011).

128. K. H. Choi *et al.*, Insight into DNA and Protein Transport in Double-Stranded DNA Viruses: The Structure of Bacteriophage N4. *J. Mol. Biol.* **378**, 726 (2008).

129. A. Fokine *et al.*, A Three-dimensional Cryo-electron Microscopy Structure of the Bacteriophage φKZ Head. *J. Mol. Biol.* **352**, 117 (2005).

130. W. Jiang *et al.*, Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat. Struct. Biol.* **10**, 131 (2003).

131. F. Yang *et al.*, Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nat. Struct. Biol.* **7**, 230 (2000).

132. A. Fokine *et al.*, Molecular architecture of the prolate head of bacteriophage T4. *Proc Natl Acad Sci U S A* **101**, 6003 (2004).

133. M. E. Cerritelli *et al.*, Encapsidated Conformation of Bacteriophage T7 DNA. *Cell* **91**, 271 (1997).

134. H. T. Jäälinoja *et al.*, Structure and host-cell interaction of SH1, a membrane-containing, halophilic euryarchaeal virus. *Proc Natl Acad Sci U S A* **105**, 8008 (2008).

135. M. J. van Raiij, A. Mitraki, G. Lavigne, S. Cusack, A triple beta-spiral in adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**, 935 (1999).

136. P. S. Rydman *et al.*, Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. *J. Mol. Biol.* **291**, 575 (1999).

137. N. V. Hud, K. H. Downing, Cryoelectron microscopy of λ phage DNA condensates in vitreous ice: the fine structure of DNA toroids. *Proc Natl Acad Sci U S A* **98**, 14925 (2001).

138. A. M. Grahn, R. Daugelavicius, D. H. Bamford, Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. *Mol. Microbiol.* **46**, 1199 (2002).